

Variance component estimation uncertainty for unbalanced data: application to a continent-wide vertical datum

M.S. Filmer · W.E. Featherstone ·

S.J. Claessens

Received: date / Accepted: date

M.S. Filmer

Western Australian Centre for Geodesy & The Institute for Geoscience Research,
Curtin University of Technology, GPO Box U1987, Perth, WA 6845, Australia

Tel.: +61-8-9266-2582

Fax: +61-8-9266-2703

E-mail: M.Filmer@curtin.edu.au

W.E. Featherstone

Western Australian Centre for Geodesy & The Institute for Geoscience Research,
Curtin University of Technology, GPO Box U1987, Perth, WA 6845, Australia

Tel.: +61-8-9266-2734

Fax: +61-8-9266-2703

E-mail: W.Featherstone@curtin.edu.au

S.J. Claessens

Western Australian Centre for Geodesy & The Institute for Geoscience Research,
Curtin University of Technology, GPO Box U1987, Perth, WA 6845, Australia

Tel.: +61-8-9266-3505

Fax: +61-8-9266-2703

E-mail: S.Claessens@curtin.edu.au

Abstract Variance component estimation (VCE) is used to update the stochastic model in least-squares adjustments, but the uncertainty associated with the VCE-derived weights is rarely considered. Unbalanced data is where there is an unequal number of observations in each heterogeneous data set comprising the variance component groups. As a case study using highly unbalanced data, we redefine a continent-wide vertical datum from a combined least-squares adjustment using iterative VCE and its uncertainties to update weights for each set. These are: (1) a continent-wide levelling network, (2) a model of the ocean's mean dynamic topography and mean sea level observations, and (3) GPS-derived ellipsoidal heights minus a gravimetric quasigeoid model. VCE uncertainty differs for each observation group in the highly unbalanced data, being dependent on the number of observations in each group. It also changes within each group after each VCE iteration, depending on the magnitude of change for each observation group's variances. It is recommended that VCE uncertainty is computed for VCE updates to the weight matrix for unbalanced data so that the quality of the updates for each group can be properly assessed. This is particularly important if some groups contain relatively small numbers of observations. VCE uncertainty can also be used as a threshold for ceasing iterations, as it is shown - for this data set at least - that it is not necessary to continue time consuming iterations to fully converge to unity.

Keywords Variance component estimation (VCE) · VCE uncertainty · vertical datum · combined least-squares adjustment · unbalanced data

1 Introduction

Variance component estimation (VCE) is used in many disciplines to obtain or improve estimates of the variances for heterogeneous observation types. Our interest in VCE stems from a previous investigation into the use of heterogeneous height data in a combined least-squares adjustment (CLSA) to redefine a continent-wide vertical datum, and the problems encountered in determining realistic relative weights among the heterogeneous observation types. Reviewers' comments on an earlier version of this paper recommended the use of VCE to determine weights for the different height data, leading to a suggestion from Teunissen (2012 pers. comm.) to also compute the uncertainty of the computed variance components based on the work of Amiri-Simkooei (2007) and Teunissen and Amiri-Simkooei (2008).

During this additional research, we found that although VCE is often computed for many applications (e.g., geodesy, genetics, environmental and medical studies, to name but a few), their uncertainties are rarely considered, or applied practically (cf. Koch 1999, p.273; Crocetto et al. 2000; Amiri-Simkooei et al. 2007, Amiri-Simkooei 2009; Amiri-Simkooei 2013). Knowing the uncertainty of the computed variance components provides two benefits: (1) it allows an analysis of the quality of the updated stochastic information for each observation group and how this propagates into the adjusted parameters (e.g., Amiri-Simkooei 2009), and (2) it provides a threshold for iterative VCE procedures, which can avoid running redundant time-consuming iterations until the $VCE = 1$ for all observation groups (e.g., Fotopoulos 2005). This is particularly useful in the case of unbalanced data sets, where each heterogeneous group contain different numbers of observations (e.g., Samanta and Welsh 2013), because VCE uncertainty primarily depends on the number of observations in each group.

The use of VCE and its uncertainty is applied in a case study where we demonstrate that a continent-wide vertical datum can be redefined using a CLSA of heterogeneous height data. The motivation for redefining vertical datums comes from the biases and/or regional distortions now evident in decades-old levelling-only vertical datums, caused primarily by the use of approximate or incorrect methods and data in their initial realisation (e.g., Filmer and Featherstone 2009; Featherstone and Filmer 2012; Penna et al. 2013). These include (among many other error sources) applying approximate height corrections to levelling (e.g., Filmer et al. 2010) and fixing levelling networks to mean sea level (MSL) observed at multiple tide gauges (e.g., Roelse et al., 1971). The advent of GPS, continued development of gravimetric quasi/geoid models, improved models of the ocean’s mean dynamic topography (MDT), gravity data to apply more accurate height corrections to levelling, and updated levelling data allow a CLSA of the heterogeneous height data, facilitated by increased computing power now available, so that a least-squares adjustment (LSA) of a continent-wide network can be done in a single operation.

We advocate that when sufficient new data are available, ageing vertical datums should be redefined using a combination of the above data (e.g., Filmer and Featherstone 2012b) rather than just levelling fixed to MSL, so as to produce the most accurate vertical datum permitted by the newer data. This follows earlier proposals from Kearsley et al. (1993) and Hwang (1997) and is a preferable solution to simply ‘fitting’ quasi/geoid models to old vertical datums because the vertical datum is the reference frame upon which all heights are built, but remains corrupted. For example, a poor quality vertical datum with a ‘fitted’ quasi/geoid model is then only compatible with that particular model, so that other regional and global height data (e.g., EGM2008 [Pavlis et al. 2012; 2013], high-precision satellite-derived digital eleva-

tion models (DEM), MDT models) continue to expose problems in the vertical datum (Featherstone and Filmer 2012; Penna et al. 2013; Véronneau et al. 2006; Smith and Roman 2001).

There are numerous challenges that need to be overcome in the development of a CLSA using heterogeneous data. These include (1) identifying and treating outliers; (2) identification and treatment of systematic errors and; (3) determination of appropriate *a priori* observation weights for the different data sets. We will focus on (3) in this paper, using VCE and its uncertainty in an empirical study of the problems associated with estimating a stochastic model for continent-wide vertical datum redefinition from heterogeneous data.

2 Variance component estimation

2.1 Background

Helmert (1924) developed a method for unbiased variance estimates for heterogeneous data, rather than using an overall variance factor on the assumption that the relative information within the weight matrix of observations was correct. Rao (1971) developed minimum norm quadratic unbiased estimation (MINQUE), but this has been shown to produce the same results as Helmert (1924) when the observations are normally distributed (e.g., Fotopoulos 2005), a property that is common to the numerous other VCE methods subsequently presented.

In addition to MINQUE, VCE methods include least-squares VCE (LS-VCE) (Teunissen and Amiri-Simkooei 2008) and the best invariant quadratic unbiased estimates (BIQUE; e.g., Sjöberg 1984). There are other VCE methods, plus numerous other

studies on these methods and the reader is directed to, e.g., Grafarend (1985), Searle (1995), Crocetto et al. (2000), Fotopolous (2003, 2005).

VCE computations are often very time consuming, requiring repeated matrix multiplication and inversion, so that for large data sets, the computational load required for iterative VCE tends to inhibit its widespread use. For this reason we use a simplified version of the iterative BIQUE (Section 2.2), based on Caspary (1987). We use the LS-VCE method for computing VCE uncertainty presented in Amiri-Simkooei (2007) and Teunissen and Amiri-Simkooei (2008) (Section 2.3). To test our results, and also to compare efficiencies in the computational load, we also use the rigorous LS-VCE and its VCE uncertainty from Amiri-Simkooei (2009).

The BIQUE delivers invariant and unbiased estimators, but one drawback is that when the number of observations is small or the stochastic model is incorrect, it may produce negative variance components (e.g., Sjöberg 1984), which are of no use to update the covariance matrix. The BIQUE requires some prior knowledge of the distribution of the observations (assumed to be normal).

Several different models can be used in VCE estimation (see Fotopoulos 2003, p. 121), of which the most common is the Gauss-Markov model, represented as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}; D\{\mathbf{y}\} = \mathbf{Q}_y \quad (1)$$

where \mathbf{y} is the $(m \times 1)$ vector of observables, \mathbf{A} is the $(m \times n)$ design matrix, \mathbf{x} is the $(n \times 1)$ unknown parameter vector and \mathbf{v} is the vector of residuals. $D\{\mathbf{y}\}$ is the dispersion or variance of the observation vector \mathbf{y} and \mathbf{Q}_y is its variance matrix. Our principal interest here is the composition of \mathbf{Q}_y , which contains multiple variance groups relating to different observation groups, each comprising different numbers of observations.

If we assume that \mathbf{y} is composed of k heterogeneous observation groups held in subvectors \mathbf{y}_i ($i = 1, \dots, k$), then

$$\mathbf{y} = [\mathbf{y}_1^T \mathbf{y}_2^T \dots \mathbf{y}_k^T]^T \quad (2)$$

and the stochastic information for each \mathbf{y}_i is contained in

$$\mathbf{Q}_y = \sum_{i=1}^k \hat{\sigma}_i^2 \mathbf{Q}_i \quad (3)$$

Thus, $\hat{\sigma}_i^2$ and \mathbf{Q}_i are the variance component and cofactor matrices of \mathbf{y}_i respectively. \mathbf{Q}_i and \mathbf{Q}_y are both $m \times m$ square matrices. This scheme is equivalent to that of Caspary (1987, p. 98) and is simplified in that it considers no correlation between the different observation groups' \mathbf{y}_i and their $\hat{\sigma}_i^2$, and that there is only one $\hat{\sigma}_i^2$ per \mathbf{y}_i (cf. Sjöberg 1984; Crocetto et al. 2000).

2.2 BIQUE VCE

A solution for the simplified iterative BIQUE, as suggested by Crocetto et al. (2000) and based on Caspary (1987, Ch. 8.7), is

$$\hat{\sigma}_i^2 = \frac{\hat{\mathbf{v}}_i^T \mathbf{P}_i \hat{\mathbf{v}}_i}{\text{tr}(\mathbf{Q}_{\hat{\mathbf{v}}_i} \mathbf{P}_i)} \quad (4)$$

where $\hat{\mathbf{v}}_i$ is the LS-residual, $\text{tr}()$ is the trace of the matrix in the parenthesis, $\mathbf{Q}_{\hat{\mathbf{v}}_i}$ the cofactor matrix of the LS residuals and \mathbf{P}_i the weight matrix, all of the i^{th} group of observations. As $\text{tr}(\mathbf{Q}_{\hat{\mathbf{v}}_i} \mathbf{P}_i) = \sum_{j=1}^{m_i} r_j$ (Crocetto et al. 2000), where r_j is the redundancy number of each observation j and m_i is the number of observations, both within the i^{th} observation group, the solution can be simplified further to (cf. Ou 1989, p.147)

$$\hat{\sigma}_i^2 = \frac{\left(\sum_{j=1}^{m_i} \hat{v}_j^2 P_j \right)_i}{\left(\sum_{j=1}^{m_i} r_j \right)_i} \quad (5)$$

$P_j = 1/\sigma_j^2$ is the weight of observation j , with $r_j = 1 - (\sigma_j^2/\hat{\sigma}_j^2)$ (e.g., Teunissen 2006b, p. 104), where σ_j^2 is the *a priori* and $\hat{\sigma}_j^2$ the *a posteriori* variance of observation j , all within each observation group i . Equation (5) holds only when the weight matrix is diagonal.

Herein, $\hat{\sigma}_{i_\ell}^2$ ($\ell = 0, 1, 2, \dots, F$; F denoting the final iteration) is defined as the iterated variance component (see Section 2.5), and is distinct from $\hat{\sigma}_i^2$ which is defined as the product of all iterated variance components (Crocetto et al. 2000)

$$\hat{\sigma}_i^2 = \prod_{\ell=1}^F \hat{\sigma}_{i_\ell}^2 \quad (6)$$

The iterative procedure requires the first VCE iteration ($\hat{\sigma}_{i_1}^2$) to update \mathbf{Q}_{y_0} so that

$$\mathbf{Q}_{y_1} = \sum_{i=1}^k \hat{\sigma}_{i_1}^2 \mathbf{Q}_{i_0} \quad (7)$$

with subscript 0 indicating the initial stochastic information prior to starting VCE iterations, which usually continue $\forall i$ until $\hat{\sigma}_{i_F}^2 = 1$ (Fotopoulos 2005). We will test whether this is practically necessary (Section 4.2.2), because this is an issue of some importance, considering the amount of time required to compute each iteration using a rigorous VCE method.

In most situations, *a priori* variances can be estimated empirically, so that \mathbf{Q}_{y_0} is taken to be a reasonable approximation of the true \mathbf{Q}_y . This is desirable for several reasons, but most notably because many applications require outlier detection procedures to identify and treat blunders prior to VCE and the final LSA of parameters, and also because fewer iterations are required if each \mathbf{Q}_{i_0} is reasonably close to its true value.

2.3 LS-VCE

LS-VCE is described in detail by Amiri-Simkooei (2007) and Teunissen and Amiri-Simkooei (2008), and applied principally to improve knowledge of the GPS stochastic model (see Teunissen and Amiri-Simkooei (2008) and references therein). Of particular benefit with the LS-VCE method is the ability to compute the uncertainty of the computed variance component estimate (Section 2.4). The LS-VCE stochastic model is (Teunissen and Amiri-Simkooei 2008)

$$\mathbf{Q}_y = \mathbf{Q}_0 + \sum_{i=1}^k \hat{\sigma}_i^2 \mathbf{Q}_i \quad (8)$$

In this case, we are interested in only the variance component, so use the notation $\hat{\sigma}_i^2$, but Eq. (8) can also be used to compute covariance components, where Teunissen and Amiri-Simkooei (2008) use different notation to represent covariance components. The co-factor matrices \mathbf{Q}_i are assumed known, with \mathbf{Q}_0 the known part of \mathbf{Q}_y assumed to be positive definite. When all known information is in each \mathbf{Q}_i , \mathbf{Q}_0 can be considered zero (e.g., Amiri-Simkooei 2013). When $\mathbf{Q}_0 = \mathbf{0}$ and there is no covariance information (\mathbf{Q}_i is diagonal), Eq. (3) is the same as Eq. (8) (cf. Teunissen and Amiri-Simkooei 2008, Section 7).

The $k \times 1$ vector $\hat{\sigma}^2$ containing all $\hat{\sigma}_i^2$ is computed as, e.g., Amiri-Simkooei (2009)

$$\hat{\sigma}^2 = \mathbf{N}^{-1} \mathbf{f} \quad (9)$$

where \mathbf{N} is the $k \times k$ normal matrix and \mathbf{f} is a $k \times 1$ vector. The elements of \mathbf{N} are computed as

$$n_{pq} = \frac{1}{2} \text{tr}(\mathbf{P}_y \mathbf{P}_A^\perp \mathbf{Q}_p \mathbf{P}_y \mathbf{P}_A^\perp \mathbf{Q}_q) \quad (10)$$

and for \mathbf{f}

$$f_p = \frac{1}{2} \hat{\mathbf{v}}^T \mathbf{P}_y \mathbf{Q}_p \mathbf{P}_y \hat{\mathbf{v}} \quad (11)$$

where the weight matrix $\mathbf{P}_y = \mathbf{Q}_y^{-1}$, $\hat{\mathbf{v}} = \mathbf{P}_A^\perp \mathbf{y}$, \mathbf{Q}_p and \mathbf{Q}_q represent the cofactor matrices \mathbf{Q}_i but where p is the row number ($p = 1, \dots, k$) in \mathbf{N} and q is the column number ($q = 1, \dots, k$), thus determining the \mathbf{Q}_i to be used in the computation for each element. Matrix \mathbf{P}_A^\perp projects onto a subspace orthogonal to \mathbf{A} and is computed as

$$\mathbf{P}_A^\perp = \mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{P}_y \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}_y \quad (12)$$

with \mathbf{I} being the identity matrix.

The iterative procedure based on Eq. (7) can also be applied to the LS-VCE method shown here. Amiri-Simkooei (2009) suggests that 10 iterations are usually sufficient to obtain converged variance components using LS-VCE, using 17-20 iterations to reach pre-set thresholds in Amiri-Simkooei (2013), but commenting that only a few iterations are practically required. Fotopoulos (2005) used 45 and 76 iterations for the I-AUE and I-MINQUE methods respectively, although appears to continue the iterations strictly until $\hat{\sigma}_{i\ell}^2 = 1$. Both studies use balanced data.

2.4 VCE uncertainty

Koch (1999, p. 273) and Crocetto et al. (2000) show formulas for the variance of $\hat{\sigma}_i^2$ ($\sigma_{\hat{\sigma}_i^2}^2$) computed using iterative BIQUÉ, but these do not seem to be widely applied in the literature. As an alternative, Davies and Blewitt (2000) computed Monte Carlo confidence intervals to derive the probability distribution of $\hat{\sigma}_i^2$ prior to their application.

The LS-VCE $\sigma_{\hat{\sigma}_i^2}^2$ for each group of heterogeneous height data can be computed (iteratively) by the inversion of \mathbf{N} (needed for Eq. (9)). The $k \times k$ covariance matrix of $\hat{\sigma}_i^2$ ($\mathbf{Q}_{\hat{\sigma}^2}$) is computed as (e.g., Amiri-Simkooei 2009)

$$\mathbf{Q}_{\hat{\sigma}^2} = \mathbf{N}^{-1} \quad (13)$$

where $\sigma_{\hat{\sigma}_i^2}^2$ comprises the diagonal elements of $\mathbf{Q}_{\hat{\sigma}^2}$. If a simplified VCE (SVCE) method such as Eq. (5) is used, an alternative LS-VCE method can be used to compute $\mathbf{Q}_{\hat{\sigma}^2}$ (Teunissen and Amiri-Simkooei 2008)

$$\mathbf{Q}_{\hat{\sigma}^2} = \mathbf{H}^{-1} \mathbf{M} \mathbf{H}^{-1} \quad (14)$$

The entries for each element h_{pq} of \mathbf{H} are (for observation equation)

$$h_{pq} = \text{tr}(\mathbf{Q}_p \mathbf{P}_y \mathbf{P}_A^\perp \mathbf{Q}_q \mathbf{P}_y \mathbf{P}_A^\perp) \quad (15)$$

and for \mathbf{M}

$$\begin{aligned} m_{pq} = & 2(\kappa + 1) \text{tr}(\mathbf{Q}_p \mathbf{P}_y \mathbf{P}_A^\perp \mathbf{Q}_y \mathbf{P}_y \mathbf{P}_A^\perp \mathbf{Q}_q \mathbf{P}_y \mathbf{P}_A^\perp \mathbf{Q}_y \mathbf{P}_y \mathbf{P}_A^\perp) \\ & + \kappa \text{tr}(\mathbf{Q}_p \mathbf{P}_y \mathbf{P}_A^\perp \mathbf{Q}_y \mathbf{P}_y \mathbf{P}_A^\perp) \text{tr}(\mathbf{Q}_q \mathbf{P}_y \mathbf{P}_A^\perp \mathbf{Q}_y \mathbf{P}_y \mathbf{P}_A^\perp) \end{aligned} \quad (16)$$

where p and q are as previously described. κ is the kurtosis parameter, which should be set to zero when \mathbf{y} is normally distributed (Teunissen and Amiri-Simkooei 2008). For a Student distribution where the degrees of freedom are sufficiently large, κ may be set to zero as an approximation (*ibid.*). It is assumed that Eq. (13) and Eq. (14) produce the same result, but this will be tested later in Section 4.2.2.

2.5 A remark on VCE uncertainty

Equation (5) or Eq. (9) are recomputed numerous times, updating \mathbf{Q}_y at each iteration (Eq. 7), so the output is $\hat{\sigma}_{i_\ell}^2$, from which $\hat{\sigma}_i^2$ is computed using Eq. (6). It then follows that each computation of $\sigma_{\hat{\sigma}_i^2}^2$ from Eq. (13) and Eq. (14) refers to a specific $\hat{\sigma}_{i_\ell}^2$, and is denoted $\sigma_{\hat{\sigma}_i^2}^2$. The final updated *a priori* standard deviation (SD; σ_i) is $\sigma_{i_F} = \sqrt{\hat{\sigma}_{i_F}^2} \sigma_{i_{F-1}}$ and is accepted as the *a priori* SD used in the final CLSA (Section 3). $\sigma_{i_{F-1}}$ is the penultimate SD iteration and σ_{i_0} the initial SD estimate, all for the i^{th}

observation group. Thus, we can say (cf. Eq. (6))

$$\hat{\sigma}_i^2 = \frac{\sigma_{i_F}^2}{\sigma_{i_0}^2} = \frac{\hat{\sigma}_{i_F}^2 \sigma_{i_{F-1}}^2}{\sigma_{i_0}^2} = \prod_{\ell=1}^F \hat{\sigma}_{i_\ell}^2 \quad (17)$$

so that the variance of $\hat{\sigma}_i^2$ ($\sigma_{\hat{\sigma}_i^2}^2$) following $\ell = F$ iterations can be computed as

$$\sigma_{\hat{\sigma}_i^2}^2 = \left(\frac{\sigma_{i_{F-1}}^2}{\sigma_{i_0}^2} \right)^2 \sigma_{\hat{\sigma}_{i_F}^2}^2 \quad (18)$$

where $\sigma_{\hat{\sigma}_{i_F}^2}^2$ is the final iteration of $\sigma_{\hat{\sigma}_{i_\ell}^2}^2$.

Hence, we can apply non-linear error propagation to derive the SD of the final updated SD σ_{i_F} ($\sigma_{\sigma_{i_F}}$) as

$$\sigma_{\sigma_{i_F}} = \sqrt{\left(\frac{\partial \sigma_{i_F}}{\partial \hat{\sigma}_i^2} \right)^2 \sigma_{\hat{\sigma}_i^2}^2} = \pm \frac{\sigma_{i_0} \sigma_{\hat{\sigma}_i^2}}{2\sqrt{\hat{\sigma}_i^2}} \quad (19)$$

Equation (19) is comparable to the result in Amiri-Simkooei (2009, Appendix A) and is used in Section 4.2.2 to compute the contribution of $\sigma_{\hat{\sigma}_i^2}^2$ to the error in σ_{i_F} . This provides an indication of how $\sigma_{\hat{\sigma}_i^2}^2$ propagates into the final adjusted heights, although strictly, this cannot be done as it would be a deterministic application of a stochastic process.

3 Combined least-squares adjustment

3.1 Input data

The steps to compute a CLSA for a redefined vertical datum from unbalanced heterogeneous data are set out in Fig. 1. The input data required are: (1) observed height differences from a continent-wide levelling network with normal corrections (Molodensky et al. 1962) applied (ΔH^N), with gravity data on the Earth's surface required to compute height corrections, (2) GPS ellipsoid heights (h) processed in a consistent reference frame, (3) height anomalies (ζ) interpolated from a gravimetric quasigeoid

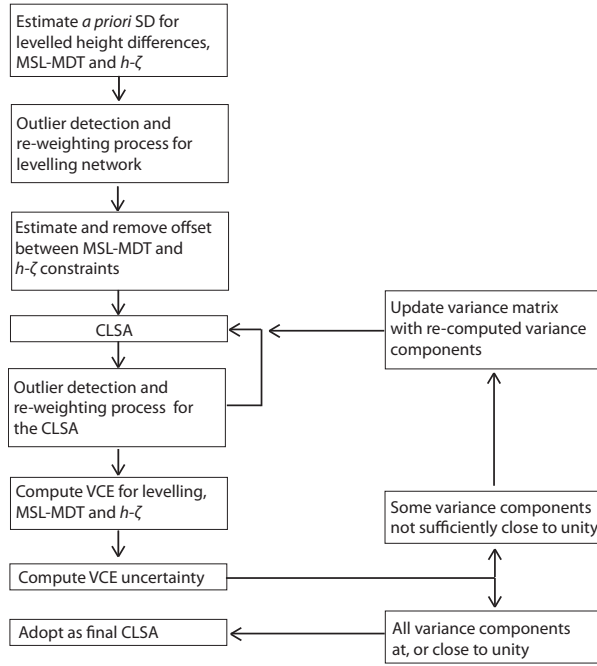


Fig. 1 Flow chart of the CLSA process.

model (4) MSL observations at tide gauges connected to the levelling network, (5) values from a model of the ocean's MDT at these tide gauges. The CLSA is arranged so that $h - \zeta$ ($H_{h\zeta}$) and MSL-MDT (H_{TG}) (both co-located with, or connected to benchmarks within the levelling network) are each treated as one observation, as per ΔH^N (cf. Fotopoulos 2005; Kearsley et al. 1993).

3.2 Method

Assuming the Gauss-Markov model in Eq. (1), the observation equation for differential levelling is

$$\Delta H_{AB}^N = \hat{H}_B^N - \hat{H}_A^N - \hat{v}_{AB} \quad (20)$$

where ΔH_{AB}^N is the levelled height difference between benchmarks A and B with the normal height correction applied, \hat{H}_B^N and \hat{H}_A^N are the LS-adjusted normal heights

of benchmarks B and A respectively, and \hat{v}_{AB} is the LS-residual of ΔH_{AB}^N . $H_{h\zeta}$ and H_{TG} constraints are treated as observations in the CLSA, so that their observation equations are

$$H_{h\zeta} = \hat{H}_{h\zeta} - \hat{v}_{h\zeta} \quad (21)$$

and

$$H_{TG} = \hat{H}_{TG} - \hat{v}_{TG} \quad (22)$$

where the constraint cofactor matrices contain realistic stochastic information, rather than very small variances designed to ‘fix’ the CLSA on the assumption that the constraints are errorless. $\hat{H}_{h\zeta}$ and \hat{H}_{TG} are the LS-adjusted normal heights for the tide gauge and $h - \zeta$ constraints respectively. All observations are held in \mathbf{y} , with the subvectors arranged as per Eq. (2), with different subvectors \mathbf{y}_i containing each of ΔH^N , $H_{h\zeta}$ and H_{TG} . If ΔH^N comprises multiple levelling types of different precision (e.g., first-order, second-order etc.), then these are arranged in additional separate subvectors. The solution to the LS-estimated heights $\hat{\mathbf{H}}$ is then

$$\hat{\mathbf{H}} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y} \quad (23)$$

This CLSA formulation provides a framework for adjusting the observations in a continental levelling network, but with the added contribution from MSL observations plus a MDT model and GPS $h - \zeta$ without ‘fixing’ these on the incorrect assumption that they are errorless. It also permits commercially available or public-domain LSA software packages to be used, thus making the process more accessible. Hwang (1997) proposes a similar CLSA, but using only MSL-MDT values as weighted constraints. Kearsley et al. (1993) uses Δh and $\Delta \zeta$ in the condition equations as height differences rather than discrete values, as in Eqs. (21) and (22). Using $\Delta \zeta$ is of no advantage, because it is simply the difference between two discrete ζ values, and not an observation.

Our scheme also differs to Kotsakis and Sideris (1999) or Fotopolous (2005) where the adjusted H (at junction point (JP) benchmarks) from levelling was used rather than the levelling observation ΔH_{AB} . In these cases, the purpose of the CLSA was to obtain systematic errors and biases through a parametric adjustment based on the condition $h - H - \zeta = 0$. Although an adjusted H can be computed from the equations in Kotsakis and Sideris (1999) and Fotopolous (2005), this is a ‘best fit’ solution, designed to enforce the $h - H - \zeta = 0$ condition and can only be realised at co-located GPS/levelling benchmarks. We use GPS $h - \zeta$ plus MSL-MDT as an additional height constraint for the levelling observations, as proposed - but not implemented - by Kearsley et al. (1993). This realises new H^N in the redefined vertical datum at all JPs in the levelling network. A further advantage of our alternative method is that it permits the estimation of $\hat{\sigma}_i^2$ for each levelling observation group, which is not possible when benchmark H are used in the adjustment in place of observed ΔH .

4 Case study: CLSA using Australian height data

We use heterogeneous height data to test the CLSA process presented (Fig. 1), with specific focus on the variance component information and its uncertainty. The outcome is an experimental continent-wide vertical datum but which does not supersede the official Australian Height Datum (AHD; Roelse et al. 1971). The AHD was realised in 1971 from a series of staged adjustments of the then Australian Levelling Survey (now the Australian National Levelling Network; ANLN) held fixed at MSL = zero at 30 mainland tide gauges. The Tasmanian AHD was realised in 1983 from the Tasmanian component of the ANLN fixed at MSL = zero at two tide gauges.

4.1 Data

The levelling data (Fig. 2) used is the ANLN (provided by Geoscience Australia [GA]; G. Johnston 2007, pers. comm.), which has received some updates since 1971. The ANLN comprises ΔH from six different levelling orders (see Table 1 in Section 4.2.2 for the number of observations in each). Third-order levelling (maximum allowable misclosure of $12\sqrt{d}$ mm, where d is the distance in km between benchmarks along the levelling route) is the dominant levelling standard in the ANLN (Roelse et al. 1971; Filmer and Featherstone 2009).

σ_{i_0} for the different levelling types were based on the estimates derived empirically from 1366 ANLN loop closures in Filmer and Featherstone (2009). These unit σ_{i_0} were propagated along each levelling section as $\sigma_{i_0}\sqrt{d}$. The estimated σ_{i_0} from Filmer and Featherstone (2009; $\text{FF}\sigma_{i_0}$ in Table 2) were increased slightly (σ_{i_0} in Table 2) to allow for the likelihood of compensating errors in the ANLN that would make the empirical estimates over-optimistic and not properly reflect the quality of the levelling (cf. Section 4.2.2).

An iterative outlier detection process (Fig. 1) was undertaken (e.g., Schwarz and Kok 1993) on the ANLN using repeated minimal constraints LSA (fixed at one mainland tide gauge and one Tasmanian tide gauge) to identify and re-weight levelling blunders that remain in the ANLN (re-observation was not possible). A significance level of $\alpha = 0.001$ was used, so that the critical value (CV) for the w -test is ± 3.29 ($w_j = \hat{v}_j/\sigma_{\hat{v}_j}$; Teunissen 2006b, p. 134). Rather than using an established method of re-weighting outliers (e.g., the Danish method) we determined the re-weighting based on adjacent ANLN loop misclosures, and the minimal detectable bias (MDB; see Teunissen 2006a, p. 67) for each observation using power level $\gamma_P = 0.80$. This was because

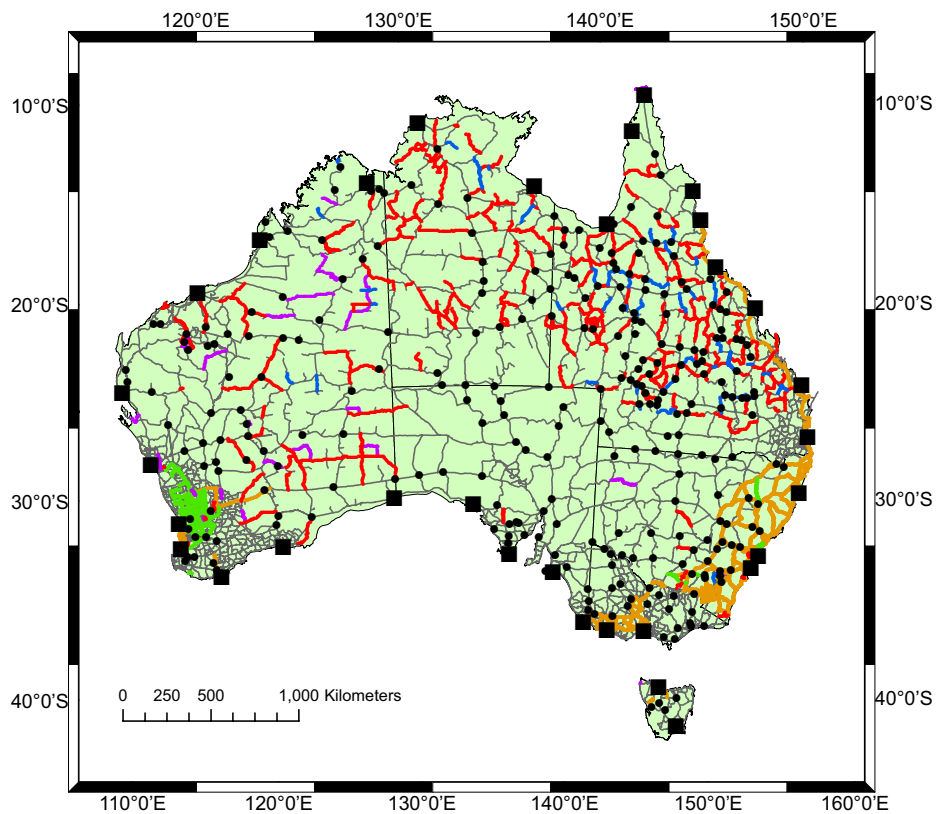


Fig. 2 The Australian national levelling network (ANLN). First order sections are in *orange*, second order in *green*, third order in *grey*, fourth order in *purple*, one-way (third order) in *red* and two-way (order undefined) in *blue*. The 32 tide-gauges used to fix the AHD and also to constrain the CLSA for this study are shown as black squares. The 277 GPS points used are black circles. Lambert projection, ANLN, GPS and tide gauge data courtesy of Geoscience Australia.

the low redundancy and suspected multiple outliers in some remote parts of the ANLN required manual assessment to reduce the risk that a ‘good’ observation (not containing an error) could be incorrectly re-weighted (type I error), or one (or more) observations that may contain errors could be accepted as correct (type II error). The process was iterated until $|w_j| < |CV| \forall \Delta H^N$. We assume for the purpose of this study that all blunders were appropriately re-weighted.

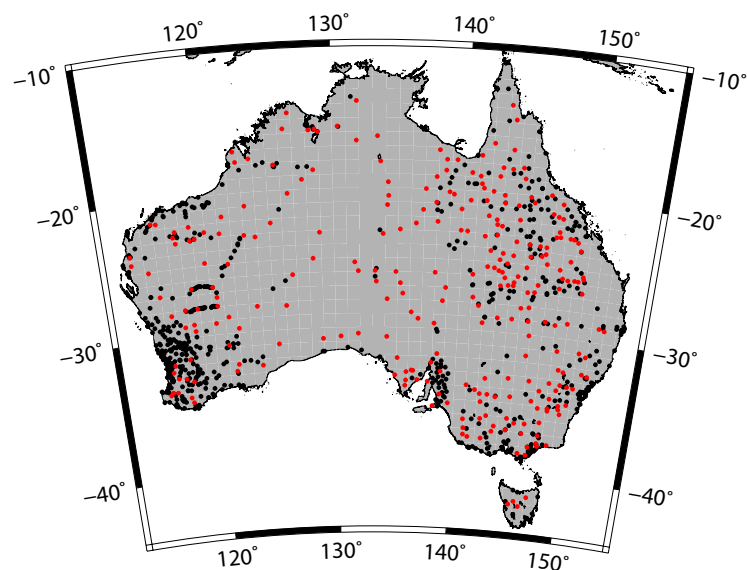


Fig. 3 The 277 GPS points used to constrain the CLSA are red circles (cf. Fig. 2) and the 765 GPS points not used are black circles. All red and black circles comprise the 1042 GPS points used to estimate the CARS2009-ITRF2000 offset. Lambert projection, GPS data courtesy of Geoscience Australia.

The $1' \times 1'$ AGQG09 quasigeoid model (Featherstone et al. 2011) is used for this study, with the AGQG09 height anomaly (ζ) bicubically interpolated at the ϕ and λ of each GPS h . AGQG09 is the gravimetric component of AUSGeoid09 (Featherstone et al. 2011). A set of 1,052 3D GPS coordinates were supplied to us by GA (N. Brown 2009, pers. comm.), processed in the International Terrestrial Reference Frame 2005 (ITRF2005; Altamimi et al. 2007) at epoch 2000 (Hu 2009). Ten GPS points were removed with apparent errors of up to 1.5 m which were assumed to be antenna height blunders. Of the 1,042 GPS points (Fig. 3), 765 were not used as constraints in the CLSA, because many were redundant observations connected to the same benchmark, and some were observed tens of kms from the nearest levelling JP to which they were connected. We selected 277 GPS points as CLSA constraints (Fig. 2 and Fig. 3), which

provided a reasonably even distribution across the ANLN, but also constraining specific parts of the ANLN based on information from the outlier detection process.

The GPS-derived h were provided with their associated SD (σ_h) from the internally propagated precision of the Bernese processing. The average σ_h for the entire set after scaling the internal precision by 10 (e.g., Rothacher 2002) is ± 26 mm (N. Brown 2009, pers. comm.), which was adopted as σ_h for all GPS points. The full variance-covariance matrix from the Bernese processing was not available. An approximate *a priori* SD estimate of ± 100 mm is made for the combined $h - \zeta$ constraint ($\sigma_{h\zeta}$) (cf. Featherstone et al. 2011).

The MSL observations used are the 32 (mostly) three-year tide gauge observations used in the realisation of the AHD (Fig. 2; Roelse et al. 1971), primarily because of availability, as many tide gauges with longer observation periods are not directly connected by levelling to the ANLN. These 32 AHD tide gauge records were used by Featherstone and Filmer (2012) to demonstrate that a combination of MSL-MDT, $h - \zeta$ at tide gauges and the ANLN removed the north-south tilt in the AHD, suggesting that they are a reasonable representation of (relative) MSL around Australia.

The MDT model used is the oceanographic-only MDT computed from the digital climatology, CSIRO Atlas of Regional Seas 2009 (CARS2009; Ridgway et al. 2002; Dunn and Ridgway 2002), which is available at <http://www.marine.csiro.au/dunn/cars2009/>. Featherstone and Filmer (2012) demonstrated that the oceanographic-only MDT model CARS2009 performed better in coastal regions than geodetic-only MDTs (satellite altimetry-derived mean sea surface (MSS) minus a geoid model) and combined MDT models (combination of geodetic and oceanographic MDTs).

An approximate *a priori* SD for combined MSL-MDT observations at 32 tide gauges (σ_{TG}) is ± 50 mm. This is estimated to comprise ± 20 mm for σ_{MSL} and $\sim \pm 45$ mm

for σ_{MDT} (cf. Filmer 2014) on the assumption that no correlation exists within \mathbf{Q}_{TG_0} . There are no formal error estimates for either of these data sets, so therefore no covariance information. We will assume \mathbf{Q}_{TG_0} to be diagonal for this reason, and also for ease of computation, as is often the case in geodetic applications, although acknowledging the impact that excluding covariance information may have on the computed VCE (e.g., Fotopoulos 2005).

A bias of -165 ± 120 mm is calculated between the mean of 1,042 $H_{h\zeta}$ and the corresponding H^N from a LSA of the ANLN constrained at 32 tide gauges by H_{TG} (CARS2009). We used the full data set (1042 GPS points) to estimate the datum bias because we considered this to be representative of the complete data set. The bias using 277 GPS points selected as CLSA constraints was -171 ± 134 mm, which can be considered the same as for the 1042 GPS points given that the 6 mm difference is much less than the associated precision. The bias is removed from CARS2009 MDT heights so that the zero-reference level for the CLSA is ITRF2005 (epoch 2000) as determined in the processing of h . Vertical datum unification between ANLN(mainland) and ANLN(Tas) is achieved through CARS2009 at 30 AHD tide-gauges on the mainland and two in Tasmania, by removing the CARS2009 MDT offset (Filmer and Featherstone 2012a).

4.2 Results

4.2.1 VCE computation

The CLSA was conducted using the Survey Network Adjustment Program (SNAP), which is freely available at <http://www.linz.govt.nz/geodetic/software-downloads>. This software allows for the CLSA formulation with the weighted constraints as set out in

Section 3.2. The output from this adjustment included the necessary information r_j , \hat{v}_j , and the input \mathbf{Q}_y . The SVCE was computed using code written to (1) separate the SNAP output and input data into the separate components; (2) compute $\hat{\sigma}_{i_\ell}^2 \forall i$ using Eq. (5); (3) scale each \mathbf{Q}_i to update \mathbf{Q}_y ; and (4) iterate this procedure until $\hat{\sigma}_{i_\ell}^2$ is (or very close to) unity. Due to the diagonal \mathbf{Q}_y and the simplified solution used, each iteration is computed in a few seconds.

The computation of $\sigma_{\hat{\sigma}_i^2}^2$ using Eqs. (14) to (16) is a much more time consuming process because each computation (iteration) requires the repeated multiplication of $\sim 7,600 \times 7,600$ matrices and several inversions of similar sized arrays. In Eq. (16), κ is set to zero (Teunissen and Amiri-Simkooei 2008), as it is assumed that our data is normally distributed. We used Crout's algorithm (Press et al. 1992, p. 33) for lower-upper decomposition for matrix inversion, which was adapted from subroutines in Press et al. (1992, Ch.2.3-2.4). Each computation (one iteration) of $\mathbf{Q}_{\hat{\sigma}^2}$ took ~ 50 hours on a Linux server (256GB RAM, 2.90GHz CPU). This code was modified to compute LS-VCE using Eqs. (9) to (13), which took ~ 23 hours per iteration, but this yielded both $\hat{\sigma}_i^2$ and $\sigma_{\hat{\sigma}_i^2}^2$.

4.2.2 VCE results

The results are arranged in Table 1 to show $\hat{\sigma}_i^2$, $\sigma_{\hat{\sigma}_i^2}^2$ and $\sigma_{\hat{\sigma}_i^2}$ computed using Eq. (6) and Eq. (18) respectively, with $\sigma_{\hat{\sigma}_i^2} = \sqrt{\sigma_{\hat{\sigma}_i^2}^2}$. Table 2 shows σ_{i_ℓ} representing the SD of the different observation types after each VCE iteration, with $\sigma_{\hat{\sigma}_{i_F}}$ ($F = 7$) computed as per Eq. (19), which indicates the contribution of $\sigma_{\hat{\sigma}_i^2}^2$ to the uncertainty in σ_{i_7} . Graphical representations of $\hat{\sigma}_{i_\ell}^2$ and $\sigma_{\hat{\sigma}_{i_\ell}^2}$ in Figs. 4 - 8 show comparisons of the LS-VCE and SVCE converging towards unity, and the changes in $\sigma_{\hat{\sigma}_{i_\ell}^2}$ with each iteration. All $\hat{\sigma}_i^2$ were positive, but the off diagonal elements in $\mathbf{Q}_{\hat{\sigma}^2}$ were small and mostly

Table 1 $\hat{\sigma}_i^2$, $\sigma_{\hat{\sigma}_i^2}^2$ and $\sigma_{\hat{\sigma}_i^2}$ computed for the SVCE. The number of observations for each observation group and their percentage of total number of observations (7614) is also shown.

$\hat{\sigma}_i^2$, $\sigma_{\hat{\sigma}_i^2}^2$ and $\sigma_{\hat{\sigma}_i^2}$ are unitless.

	$\hat{\sigma}_i^2$	$\sigma_{\hat{\sigma}_i^2}^2$	$\sigma_{\hat{\sigma}_i^2}$	no. obs	% of total
First	0.9852	0.0133	± 0.1155	941	12.4
Second	1.3590	0.0613	± 0.2477	282	3.7
Third	1.2145	0.0023	± 0.0474	5630	73.9
Fourth	2.1605	0.4753	± 0.6894	51	0.7
One-way	0.9897	0.0129	± 0.1135	347	4.6
Two-way	1.0961	0.1155	± 0.3399	54	0.7
MSL-MDT	2.7201	0.8319	± 0.9121	32	0.4
$h - \zeta$	0.6310	0.0048	± 0.0692	277	3.6

negative, reflecting that no covariance information was used in \mathbf{Q}_y . The SVCE and LS-VCE $\hat{\sigma}_i^2$ were the same (cf. Teunissen and Amiri-Simkooei 2008, Section 7), or very close, as shown in Table 2.

The highly unbalanced data (Table 1) has resulted in a wide range of uncertainties for $\hat{\sigma}_i^2$, indicating that for first-order, one-way and two-way levelling the difference between $\hat{\sigma}_i^2$ and unity is less than their $\sigma_{\hat{\sigma}_i^2}$. The relatively small changes to σ_{i_0} for these levelling types can also be seen in Table 2. The values for σ_{i_7} (Table 2) for each levelling order indicate that, as suggested by Filmer and Featherstone (2009), their empirical σ_i are optimistic. The reason for this is that loop-based analysis cannot account for compensating blunders or systematic errors within each loop. In addition, these levelling groups are of different quality and should not be given the same *a priori* σ_i in any LSA.

By the seventh iteration for the SVCE (Table 2), any change in σ_i is negligible. Any change after iteration four is so small that iterating past this point is difficult

Table 2 Empirically-derived (from the ANLN) σ_{i_0} from Filmer and Featherstone (2009) (FF σ_{i_0}), the σ_{i_0} adopted for the first iteration of the CLSA, and σ_{i_ℓ} following each SVCE and LS-VCE iteration. $\sigma_{\sigma_{i_7}}$ is computed for SVCE only - LS-VCE results would be the same. See Table 1 to compare the number of observations for each group to $\sigma_{\sigma_{i_7}}$. Units are mm, scaled by \sqrt{d} for the ANLN ΔH .

	FF σ_{i_0}	σ_{i_0}	σ_{i_1}	σ_{i_2}	σ_{i_3}	σ_{i_4}	σ_{i_5}	σ_{i_6}	σ_{i_7}	$\sigma_{\sigma_{i_7}}$
First (SVCE)	2.4	3.0	3.10	3.04	3.01	3.00	2.99	2.98	2.98	± 0.17
First (LS-VCE)	2.4	3.0	2.98	2.98	2.97	2.97	2.97	2.97	2.97	
Second (SVCE)	2.8	3.0	3.41	3.45	3.47	3.47	3.48	3.49	3.50	± 0.32
Second (LS-VCE)	2.8	3.0	3.46	3.47	3.48	3.48	3.49	3.49	3.50	
Third (SVCE)	4.2	4.5	4.88	4.91	4.93	4.94	4.95	4.95	4.96	± 0.10
Third (LS-VCE)	4.2	4.5	4.91	4.94	4.95	4.95	4.96	4.96	4.97	
Fourth (SVCE)	6.3	6.5	8.77	9.36	9.51	9.54	9.55	9.55	9.55	± 1.52
Fourth (LS-VCE)	6.3	6.5	9.44	9.55	9.54	9.54	9.53	9.53	9.53	
One-way (SVCE)	9.2	13.0	13.09	12.94	12.92	12.92	12.93	12.93	12.93	± 0.74
One-way (LS-VCE)	9.2	13.0	12.98	12.92	12.92	12.92	12.93	12.93	12.93	
Two-way (SVCE)	10.6	13.0	13.42	13.48	13.55	13.59	13.60	13.61	13.61	± 2.11
Two-way (LS-VCE)	10.6	13.0	13.51	13.58	13.60	13.61	13.61	13.61	13.61	
MDT (SVCE)	NA	50.0	67.11	76.00	80.00	81.40	82.33	82.42	82.46	± 13.83
MDT (LS-VCE)	NA	50.0	74.90	81.59	82.48	82.54	82.58	82.51	82.51	
$h - \zeta$ (SVCE)	NA	100.0	90.00	84.00	81.00	80.20	79.76	79.41	79.44	± 4.36
$h - \zeta$ (LS-VCE)	NA	100.0	86.59	81.38	79.88	79.39	79.24	79.21	79.17	

to justify. The effect of the small changes on the computed parameters (\hat{H}^N) for the CLSA after each SVCE iteration is demonstrated in Fig. 4, showing a maximum and minimum change in \hat{H}^N at iteration four of 2.5 mm and -2.7 mm (SD of changes is within ± 1 mm), respectively. This becomes 0.2 mm for each by iteration seven. It is suggested that maximum change in \hat{H}^N at 4,427 adjusted JP benchmarks < 3 mm (most changes much smaller) can be considered negligible, certainly in relation to the

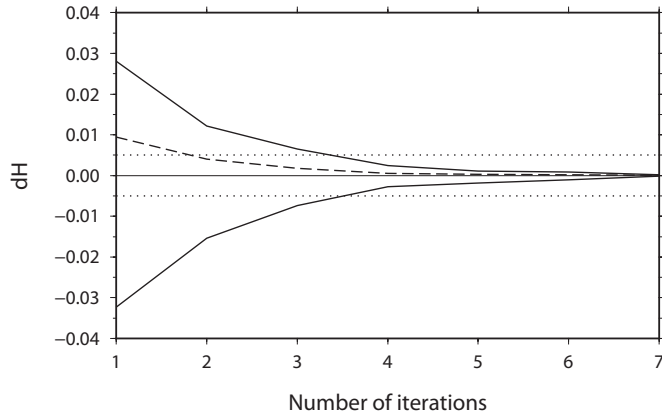


Fig. 4 Maximum, minimum (solid lines) and RMS (dashed line) of changes for all 4,427 ANLN \hat{H}^N (dH in m) after each SVCE iteration, where \mathbf{Q}_y is updated and the CLSA re-run. The dotted line is an arbitrary threshold of ± 5 mm.

precision of most of the levelling. The LS-VCE (Table 2) produces the same results, but σ_{i_ℓ} converges to unity quicker than for the SVCE (cf. Koch 1999, p. 272), although - as found by Fotopoulos (2005) - this is variable between the different observation groups (see Figs. 5 to 8).

The SD of σ_{i_τ} ($\sigma_{\sigma_{i_\tau}}$; Eq. (19)) in Table 2 indicates the contribution of the VCE uncertainty into the final adjusted parameters \hat{H}_N . This is separate from any other error components that may propagate into \hat{H}_N from the different data used and computation methods. Large $\sigma_{\sigma_{i_\tau}}$ in Table 2 are correlated with large $\sigma_{\hat{\sigma}_i^2}$ in Table 1 and the low numbers of observations for these groups. We have not estimated the possible effect of $\sigma_{\hat{\sigma}_i^2}$ on \hat{H}_N because this would be propagating a stochastic value using a deterministic process. A crude estimate of this effect can be seen by determining the iteration at which the difference between σ_{i_ℓ} and σ_{i_τ} becomes less than $\sigma_{\sigma_{i_\tau}}$. For most observation groups, this is between iterations one and two - which can also be seen in

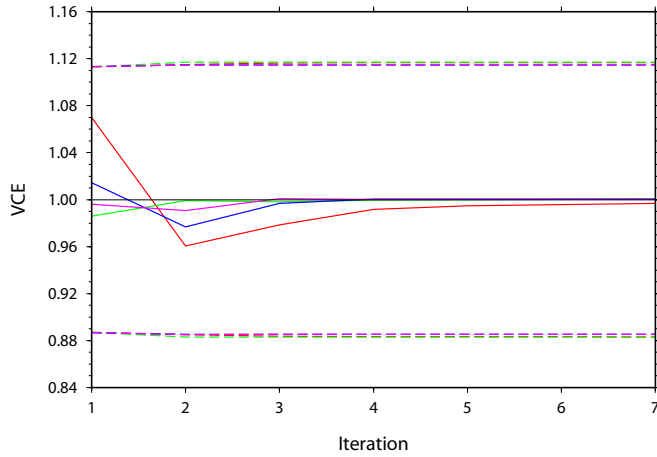


Fig. 5 Iterated $\hat{\sigma}_{i_\ell}^2$ (solid lines) for ANLN first-order (941 observations) and one-way (347 observations) levelling types with their $\sigma_{\hat{\sigma}_{i_\ell}^2}$ (dashed lines) at each iteration. First-order is green (LS-VCE) and red (SVCE), and one-way is magenta (LS-VCE) and blue (SVCE).

Fig. 5 to Fig. 8 where $\hat{\sigma}_{i_\ell}^2$ becomes less than $\sigma_{\hat{\sigma}_{i_\ell}^2}$ - so that Fig. 4 suggests that VCE uncertainty may cause an error in \hat{H}^N of up to ~ 20 -30 mm.

Figure 5 shows first-order (941 observations) and one-way (347 observations) levelling. SVCE (Eq. 5) and LS-VCE (Eq. 9) $\hat{\sigma}_{i_\ell}^2$ at each iteration are compared, as are $\sigma_{\hat{\sigma}_{i_\ell}^2}$ from LS-VCE (Eq. 13), and also from Eq. (14), which is used with the SVCE. First-order LS-VCE shows quite different behaviour to the SVCE, with the LS-VCE converging to unity after the first iteration, while SVCE struggles to fully converge to unity, although by iteration four it is sufficiently close to unity to stop iterating. Both $\sigma_{\hat{\sigma}_{i_\ell}^2}$ from Eqs. (13) and (14) are the same. One-way levelling shows similar behaviour for LS-VCE and SVCE, and although the SVCE $\hat{\sigma}_{i_\ell}^2$ over- then under-estimates more than the LS-VCE, neither exceed their $\sigma_{\hat{\sigma}_{i_\ell}^2}$ with both converging to unity by the fourth iteration.

Second- and fourth-order levelling behave similarly (Fig. 6), converging to unity by the fourth iteration, although the SVCE for fourth-order is slower. The notable change

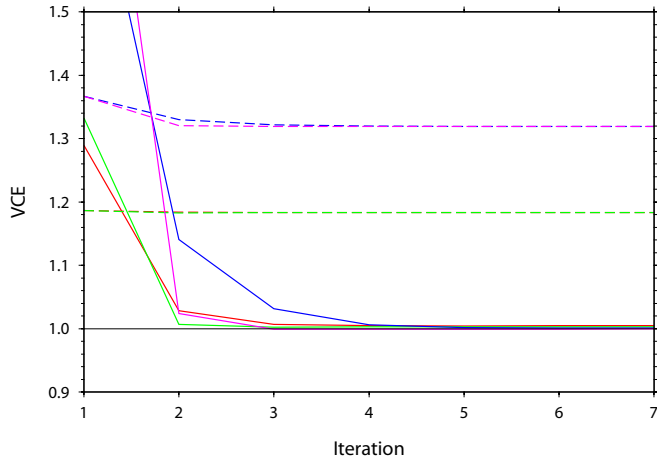


Fig. 6 Iterated $\hat{\sigma}_{i_\ell}^2$ (solid lines) for ANLN second- (282 observations) and fourth-order (51 observations) levelling types with their $\sigma_{\hat{\sigma}_{i_\ell}^2}$ (dashed lines) at each iteration. Second-order is green (LS-VCE) and red (SVCE), and fourth-order is magenta (LS-VCE) and blue (SVCE).

in $\sigma_{\hat{\sigma}_{i_\ell}^2}$ for fourth-order levelling after the first iteration suggests that VCE uncertainty should be computed for at least the first two iterations, although this is dependent on the magnitude of the first iteration (e.g., 2.107 for fourth-order LS-VCE). Figure 7 shows that third-order levelling is within its $\sigma_{\hat{\sigma}_{i_\ell}^2}$ by the second iteration, with the SVCE and LS-VCE methods providing very similar results. The two-way levelling is well within its $\sigma_{\hat{\sigma}_{i_\ell}^2}$ at iteration one, suggesting that its σ_{i_0} is a close estimate to the ‘true’ σ_i , but this is also because of large $\sigma_{\hat{\sigma}_i^2}$ which is primarily determined by the small number of observations (54) for this group (cf. fourth-order levelling with 51 observations in Fig. 6). This contrasts to the lower $\sigma_{\hat{\sigma}_i^2}$ for the third-order levelling group of 5630 observations.

H_{TG} LS-VCE $\hat{\sigma}_{i_\ell}^2$ (Fig. 8) converges by the fourth iteration, while the SVCE $\hat{\sigma}_{i_\ell}^2$ takes until its sixth iteration. The large $\sigma_{\hat{\sigma}_i^2}$ for the H_{TG} observation group (32 observations), demonstrates the uncertainty in the $\hat{\sigma}_i^2$ for this observation group (Table 2). The small difference between the two VCE uncertainty methods in H_{TG} $\sigma_{\hat{\sigma}_{i_\ell}^2}$ at

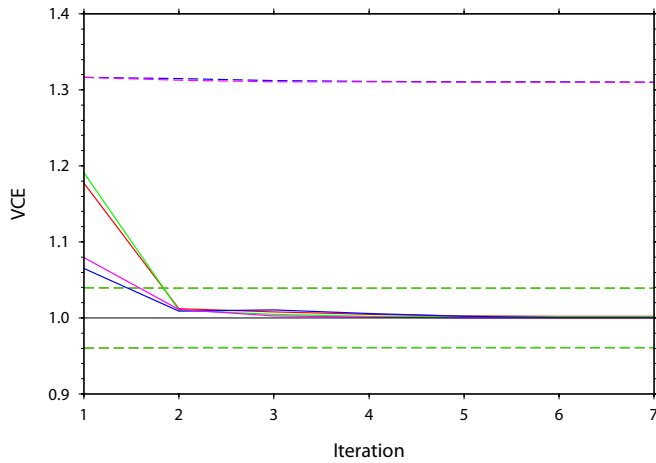


Fig. 7 Iterated $\hat{\sigma}_{i_\ell}^2$ (solid lines) for ANLN third-order (5630 observations) and two-way (54 observations) levelling types with their $\sigma_{\hat{\sigma}_{i_\ell}^2}$ (dashed lines) at each iteration. Third-order is green (LS-VCE) and red (SVCE), and two-way is magenta (LS-VCE) and blue (SVCE).

iteration two can be attributed to the different H_{TG} input \mathbf{Q}_{i_1} after updates from $\hat{\sigma}_{i_\ell}^2$ because the number of observations does not change. The change in $\sigma_{\hat{\sigma}_{i_\ell}^2}$ from iterations one to two suggests that at least two iterations of $\sigma_{\hat{\sigma}_{i_\ell}^2}$ (computed from the *a priori* \mathbf{Q}_y) should be computed. $H_{h\zeta}$ constraints converge by about iteration four, with LS-VCE only slightly faster than SVCE. There is a slight increase in $\sigma_{\hat{\sigma}_{i_\ell}^2}$ for $H_{h\zeta}$ after iteration one, but this is very small.

4.2.3 Discussion

The need to compute $\sigma_{\hat{\sigma}_i^2}$ with VCE to update the stochastic model has been demonstrated, but the exact method to use is not so clear. LS-VCE computes both $\hat{\sigma}_{i_\ell}^2$ and $\sigma_{\hat{\sigma}_{i_\ell}^2}$ for each iteration, but takes approximately one day for each iteration, depending on the power of the computer. By comparison, computing time for the (BIQUE) SVCE is a matter of seconds, so that the time to complete seven iterations depends mostly on the workflow through the LS-adjustment package to re-run each CLSA. Using the

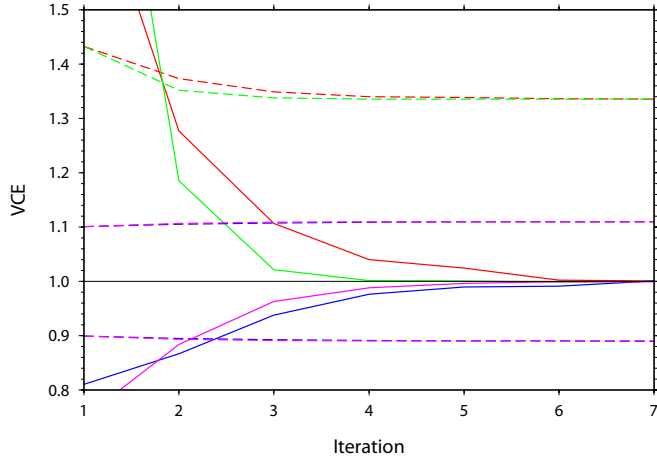


Fig. 8 Iterated $\hat{\sigma}_{i_\ell}^2$ (solid lines) for H_{TG} (32 observations) and $H_{h\zeta}$ (277 observations) constraints with their $\sigma_{\hat{\sigma}_{i_\ell}^2}$ (dashed lines) at each iteration. H_{TG} is green (LS-VCE) and red (SVCE), and $H_{h\zeta}$ is magenta (LS-VCE) and blue (SVCE).

SVCE requires the need to use Eq. (14) to compute VCE uncertainty, which takes approximately two days per iteration (again depending on computer power), or, alternatively, the full LS-VCE can be used only to obtain $\mathbf{Q}_{\hat{\sigma}^2}$, but in half the computing time. The SVCE can only be used with diagonal \mathbf{Q}_i , whereas LS-VCE can also include covariance components. When \mathbf{Q}_i are diagonal, as is often the case, SVCE can be used to compute $\hat{\sigma}_{i_\ell}^2$ with two runs of Eq. (14) (or Eq. (8)) to compute $\sigma_{\hat{\sigma}_{i_\ell}^2}$ for the first two iterations, which can also be used as a guide as to when continued iterations are unnecessary. A final run can be made to compute $\sigma_{\hat{\sigma}_{i_\ell}^2}$ for the last iteration, which is required to compute $\sigma_{\hat{\sigma}_{i_F}^2}$ using Eq. (18).

The results of the CLSA process is a set of \hat{H}^N at 4,247 ANLN JP benchmarks (not shown). The most significant feature is the removal of the north-south AHD slope (cf. Featherstone and Filmer 2012), although the additions of new levelling since 1971 have reduced some regional distortions (cf. Featherstone and Filmer 2009). In the absence of ‘true’ values to test the CLSA, a quasi-independent comparison was made

between \hat{H}^N and $H_{h\zeta}$ at 765 GPS points (Fig. 3) that were not used as constraints in the CLSA. These indicate differences that are close to a normal distribution, with the SD of these differences ± 0.094 m, which could be cautiously adopted as a proxy for external accuracy of the CLSA. The average SD for each \hat{H}^N is ± 41 mm which can be used as an internal precision for the CLSA propagated through the network from the stochastic information.

During the review process of this second submission, one reviewer suggested that because some countries may have only a few operational tide gauges with reliable MSL+MDT connected to their levelling network and/or GPS points with $h - \zeta$, there may be a lower limit (on the number of these observations) below which the CLSA and VCE method may not be viable. This is a valid point, and although we did not conduct tests to determine a minimum number, it may be a question that warrants further investigation. However, the geography and availability, accuracy and geometry of each country's geodetic data are invariably peculiar to those countries and would be impossible for us to artificially recreate. Any attempt by us to emulate these data would still be Australian-specific, simply using different combinations of the Australian data. Moreover, we suggest it is up to anyone wishing to use this method to adapt it on a case-by-case basis to their particular needs.

5 Summary and conclusions

We have demonstrated that a CLSA using unbalanced heterogeneous height data can be used to redefine a vertical datum. During this process, we have focussed on the use of VCE to improve the stochastic information for the CLSA, and investigated the need to compute VCE uncertainty for highly unbalanced data. Two different models were used

within the LS-VCE scheme. One realises both the VCE and its uncertainty, while the other computes only VCE uncertainty, taking twice (~ 1 day v ~ 2 days) the computing time as the first. The use of the BIQUÉ SVCE method provides the same result as LS-VCE for variance components (only using diagonal cofactor matrices), but permits computation in a few seconds for each iteration. One of the LS-VCE methods must then be used to compute VCE uncertainty, but it appears that - although this changes with each iteration - the changes are very small after the first or second iterations, dependent on changes to the cofactor matrix and the number of observations in the observation group. The quicker convergence of LS-VCE does not outweigh its one day computation time in comparison to the SVCE.

Computing VCE uncertainty provides information on how many iterations are needed before VCE updates are of no further practical benefit, although this depends on the desired precision. For the CLSA height data used here, there is little benefit in using more than four iterations for the SVCE. The uncertainty of variance components computed from unbalanced data are variable between groups and are highly dependent on the number of observations in each group (cf. Tables 1 and 2). It is therefore recommended that computation of VCE uncertainty become more commonplace in VCE analysis for highly unbalanced data sets, and standard when some groups contain only a few observations (e.g., <100).

Acknowledgements We would like to thank two anonymous reviewers of an earlier version of this paper, who steered us toward VCE for this study and subsequently to Peter Teunissen who suggested using LS-VCE uncertainty. Mick Filmer has received financial support from an Australian Postgraduate Award, The Institute for Geoscience Research (TIGeR), the Cooperative Research Centre for Spatial Information and the Australian Research Council (ARC) Linkage Project (project LP110100284). We would like to thank Geoscience Australia for supplying

the ANLN and GPS datasets and CSIRO Marine Laboratories for CARS2009. We appreciate the constructive comments from the Associate Editor Chris Jekeli, and reviewers; Christopher Kotsakis and two anonymous reviewers. Figures 3-8 were produced using Generic Mapping Tools (GMT; Wessel and Smith, 1998).

References

- Altamimi Z, Collileux X, Legrand J, Garayt B, Boucher C (2007) ITRF2005: A new release of the International Terrestrial Reference Frame based on time series of station positions and Earth Orientation Parameters. *Journal of Geophysical Research*, 112:B09401. doi: 10.1029/2007JB004949
- Amiri-Simkooei (2007) Least-squares variance component estimation: Theory and GPS applications. PhD Dissertation, Delft University of Technology, Publication on Geodesy, 64, Netherlands Geodetic Commission, Delft
- Amiri-Simkooei AR, Tiberius CCJM, Teunissen PJG (2007) Assessment of noise in GPS coordinate time series: methodology and results. *J Geophys Res* 112:B07413. doi:10.1029/2006JB004913
- Amiri-Simkooei AR (2009) Noise in multivariate GPS position time-series. *Journal of Geodesy*, 83(2):935-944
- Amiri-Simkooei AR (2013) Application of least squares variance component estimation to errors-in-variables models. *Journal of Geodesy*, 87(10-12):935-944
- Caspary WF (1987) Concepts of network and deformation analysis. School of Surveying, The University of New South Wales, Kensington, pp 97-111
- Crocetto N, Gatti M, Russo P (2000) Simplified formulae for the BIQUE estimation of variance components in disjunctive observation groups. *Journal of Geodesy*, 74(6):447-457

- Davies P, Blewitt G (2000) Methodology for global geodetic time series estimation: A new tool for geodynamics. *Journal of Geophysical Research*, 105(B5):11,083-11,100
- Dunn JR, Ridgway KR (2002) Mapping ocean properties in regions of complex topography. *Deep-Sea Research I*, 49(3):591:604. doi:10.1016/S0967-0637(01)00069-3
- Featherstone WE, Filmer MS (2012) The north-south tilt in the Australian Height Datum is explained by the ocean's mean dynamic topography. *Journal of Geophysical Research - Oceans*, 117:C08035
- Featherstone WE, Kirby JF, Hirt C, Filmer MS, Claessens SJ, Brown N, Hu G, Johnston GM (2011) The AUSGeoid09 model of the Australian Height Datum. *Journal of Geodesy*, 85(3):133-150. doi:10.1007/s00190-010-0422-2
- Filmer MS (2014) Using models of the oceans mean dynamic topography to identify errors in coastal geodetic levelling. *Marine Geodesy*, 37(1):47-64, doi: 10.1080/01490419.2013.868383
- Filmer MS, Featherstone WE (2009) Detecting spirit-levelling errors in the AHD: recent findings and some issues for any new Australian height datum. *Australian Journal of Earth Sciences*, 56(4):559-569 doi:10.1080/08120090902806305
- Filmer MS, Featherstone WE, Kuhn M (2010) The effect of EGM2008-based normal, normal-orthometric and Helmert orthometric height systems on the Australian levelling network. *Journal of Geodesy*, 84(8):501-513. doi:10.1007/s00190-010-0388-0
- Filmer MS, Featherstone WE (2012a) A re-evaluation of the Australian Height Datum offset between mainland Australia and Tasmania. *Marine Geodesy*, 35(1):1-13, doi:10.1080/01490419.2011.634961
- Filmer MS, Featherstone WE (2012b) Three viable options for a new Australian vertical datum. *Journal of Spatial Science*, 57(1):19-36, doi:10.1080/14498596.2012.679248

-
- Fotopoulos G (2003) An analysis on the optimal combination of geoid, orthometric and ellipsoid height data. PhD Thesis, University of Calgary, Department of Geomatics Engineering, Report No. 20185
- Fotopoulos G (2005) Calibration of geoid error models via a combined adjustment of ellipsoidal, orthometric and gravimetric geoid height data. *Journal of Geodesy*, 79(1-3):111-123. doi:10.1007/s00190-005-0449-y
- Grafarend EW (1985) Variance-covariance component estimation theoretical results and geodetic applications. *Statistics and Decisions*, 2:407-441
- Helmert FR (1924) *Die Ausgleichsrechnung nach der Methode der Kleinsten Quadrate*, 3. AUFL., Leipzig/Berlin.
- Hu G (2009) Analysis of Regional GPS campaigns and their alignment to the International Terrestrial Reference Frame (ITRF). *Journal of Spatial Science*, 54(1):15-22
- Hwang C (1997) Height system of Taiwan from satellite and terrestrial data. *Journal of Surveying Engineering*, 123(4):162-180. doi:10.1061/(ASCE)0733-9453
- Kearsley AHW, Ahmad Z, Chan A (1993) National height datums, levelling, GPS heights and geoids. *Australian Journal of Geodesy, Photogrammetry and Surveying* 59:53-88
- Koch K-R (1999) *Parameter estimation and hypothesis testing in linear models*, 2nd edn. Springer, Berlin Heidelberg Germany
- Kotsakis C, Sideris MG (1999) On the adjustment of combined GPS/levelling/geoid networks. *Journal of Geodesy*, 73(8):412-421. doi:10.1007/s001900050261
- Molodensky M, Yeremeyev V, Yurkina M (1962) *Methods for study of the external gravitational field and figure of the Earth*. Israeli Program for Scientific Translations, Jerusalem

- Ou Z (1989) Estimation of variance and covariance components. *Bulletin Géodésique*, 63:139-148
- Pavlis NK, Holmes SA, Kenyon SC, Factor JK (2012) The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research - Solid Earth*, 117(4), art no. B04406
- Pavlis NK, Holmes SA, Kenyon SC, Factor JK (2013) Correction to ‘The development and evaluation of the Earth Gravitational Model 2008 (EGM2008)’. *Journal of Geophysical Research - Solid Earth*, 118(5), art no. 10.1002/jgrb.50167
- Penna NT, Featherstone WE, Gazeaux J, Bingham RJ (2013) The apparent British sea slope is caused by systematic errors in the levelling-based vertical datum. *Geophysical Journal International*, 194(2): 772-786, doi: 10.1093/gji/ggt161
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) *Numerical Recipes in Fortran77: The art of scientific computing*. Cambridge University Press, Cambridge UK
- Rao CR (1971) Estimation of variance and covariance components - MINQUE theory. *Journal of Multivariate Analysis*, 1(3):257-275. doi:10.1016/0047-259X(71)90001-7
- Ridgway KR, Dunn JR, Wilkin JL (2002) Ocean interpolation by four-dimensional weighted least squares-application to the waters around Australasia. *Journal of Atmospheric and Oceanic Technology*, 19(9):1357-1375
- Roelse A Granger HW, Graham JW (1971) The adjustment of the Australian levelling survey 1970-1971. Technical Report 12, Division of National Mapping, Canberra, Australia
- Rothacher M (2002) Estimation of station heights with GPS. In: Drewes H, Dodson A, Fortes LPS, Sánchez L, Sandoval P (eds) *Vertical reference systems*. Springer, Berlin pp. 8190

-
- Samanta M, Welsh AH (2013) Bootstrapping for highly unbalanced clustered data. *Computational Statistics and Data Analysis*, 59:70-81
- Schwarz CR, Kok JJ (1993) Blunder detection and data snooping in LS and robust adjustments. *Journal of Surveying Engineering*, 119(4):127-136. doi:10.1061/(ASCE)0733-9453
- Searle SR (1995) An overview of variance component estimation. *Metrika*, 42:215-230
- Sjöberg LE (1984) Non negative variance component estimation in the Gauss-Helmert adjustment model. *manuscripta geodaetica*, 9:247-280
- Smith DA, Roman DR (2001) GEOID99 and G99SSS: 1-arc-minute geoid models for the United States, *Journal of Geodesy* 75(9-10):469-490, doi: 10.1007/s001900100200
- Teunissen PJG (2006a) *Network quality control*. 1st ed. Delft, Vereniging voor Studie - en Studentbelangen te Delft, The Netherlands
- Teunissen PJG (2006b) *Testing theory*. 2nd ed. Delft, Vereniging voor Studie - en Studentbelangen te Delft, The Netherlands
- Teunissen PJG, Amiri-Simkooei A (2008) Least-squares variance component estimation. *Journal of Geodesy*, 82(2):65-82. doi:10.1007/s00190-007-0157
- Véronneau M, Duval R, Huang J (2006) A gravimetric geoid model as a vertical datum in Canada. *Geomatica*, 60(2):165-172
- Wessel P, Smith WHF (1998), New, improved version of generic mapping tools released. *Eos Trans. AGU*, 79(47), 579, doi:10.1029/98EO00426.