J Geod (2005) 78: 645–653
DOI 10.1007/s00190-004-0430-1

ORIGINAL ARTICLE

L.E. Sjöberg

A discussion on the approximations made in the practical implementation of the remove–compute–restore technique in regional geoid modelling

Received: 5 February 2004 / Accepted: 12 November 2004 / Published online: 26 April 2005

Abstract The remove-compute-restore (RCR) technique is the most well known method for regional gravimetric geoid determination today. Its basic theory is the first-order approximation of either Molodensky's method for quasi-geoid determination or the classical geoid modelling by Helmert's second method of condensing the topography onto the geoid. Although the basic approximate formulae do not meet today's demands for a 1-cm geoid, it is sometimes assumed that the removal of the less precise long-wavelength terrestrial gravity anomaly field from Stokes's integral by utilising a higher-order reference field represented by a more precise Earth gravity model (EGM) and the restoration of the EGM as a low-degree geoid contribution will produce a geoid model of the desired accuracy. Further improvement is achieved also by removing and restoring a residual topographic effect, which favourably smoothes the gravity anomaly to be integrated in Stokes's formula. However, it is shown here that the RCR technique fails to tune down the long-wavelength gravity signal from the terrestrial data, and the EGM actually only reduces, in a non-optimised way, the truncation error committed by limiting the Stokes integration to a small region around the computation point. Hence, in order to take full advantage of a precise EGM, especially one from new dedicated satellite gravimetry, Stokes's kernel must be modified in a suitable way to match the errors of terrestrial gravity, EGM and truncation. In addition, topographic, atmospheric and ellipsoidal effects must be carefully applied.

Keywords Regional geoid modelling · Remove–compute–restore (RCR) technique

L.E. Sjöberg
Unit of Geodesy, Royal Institute of Technology,
100 44 Stockholm, Sweden
Tel.: +46-8-790-73-30
Fax: +46-8-790-73-43
E-mail: sjoberg@geomatics.kth.se

1 Introduction

For about a decade, the International Geoid School (IGS), under the umbrella of the International Association of Geodesy (IAG), has undertaken several training courses around the world on practical geoid determination (see e.g. Sansó 1994; Sansó and Rummel 1997). The main method for regional geoid computation taught by the IGS goes under the name of the remove–compute–restore (RCR) technique. This practical technique has been extensively applied, for example in the development of several geoid models of the Nordic Geodetic Commission (NKG; Tscherning and Forsberg 1986; Forsberg 1990, 2001). As all IGS courses have employed some teachers from the NKG, the geoid modelling experience of the NKG is an essential practical ingredient in IGS communications. Other examples of applications are the regional geoid models of Europe (EGG95; Denker et al. 1996) and the US (Milbert 1996). The influence of the RCR technique on the geodetic community is also manifested by the fact that it is the only practical gravimetric geoid modelling technique described in the recent textbook of Torge (2001, Sect. 6.7.2) to combine terrestrial gravimetric data with an Earth gravity model (EGM).

The basic theory used for the RCR technique is either a first-order approximation of M.S. Molodensky's theory for quasi-geoid determination or a classical Stokesian integration by a regularisation of the topographic masses. However, for practical application, Moritz (1966, 1980) converted Molodensky's original approach to the method by analytical continuation, originating with Bjerhammar (1962, 1963, 1969). Hence, the original formulae, in practice limited to first-order approximations, were developed when the goal for geoid determination was at a precision of at least one order of magnitude less than it is today, i.e. $\sim$10 cm or worse. However, since then the use of terrestrial gravity data in a Stokes-type solution has been improved for regional geoid determination by using a higher-order reference field taken from an EGM. In these combined solutions, the EGM is primarily intended to represent the long-wavelength gravity field, while a Stokes-type integral with residual gravity computes the high-frequency signal. Pioneer users of these ideas were Vincent and Marsh (1974) and Rapp and Rummel (1975). A common name for this type of combined geoid determination using a higher-order reference field is the RCR technique, which is usually also supplemented with the removal and restoration of the high-frequency topographic effects by a digital terrain model (DTM).

The RCR technique, as for any other method of estimating the geoid from gravimetric data, necessitates a number of corrections to the original Stokes formula. Martinec and Vaníček (1994a); Martinec and Vaníč ek (1994b) and Sjöberg (2000, 2001) have pointed to several problems in the practical evaluation of these corrections to meet the demands of a 1-cm geoid. It could be thought that the removal and subsequent restoration of a higher-order reference field and a DTM in the RCR method would justify the use of the less accurate basic formulae and corrections. This thought stems from the fact that in the RCR technique Stokes's integral does not operate on the full gravity anomaly, but only on a residual gravity anomaly reduced by the EGM and DTM. Hence, as this residual gravity anomaly signal is generally much smaller than the signal of the original gravity anomaly, it might be reasonable to believe that the RCR method can be used with cruder corrections than the original Stokes formula. The study of the validity of this idea is one major aim of this article.

The theme of this article can also be described as follows. Forsberg (2001) reported that the NKG geoid modelling by the application of the RCR technique did not improve by including more and improved gravity data over the region and surrounding areas. Consequently, if better data are not the key issue, it is likely that the approximations applied to the basic theory limit the results. Sjöberg and Ågren (2002) studied some of these issues for the NKG models, and this investigation extends that work to the RCR technique in full.

2 The basic equations of the RCR technique

As already stated, the RCR technique has its roots in the methods for quasi-geoid determination developed by Molodensky et al. (1962) and modified by Moritz (1966, 1980). Using the formulation by analytical continuation of Moritz (1980, Ch. 48), the quasi-geoidal height ζ at the surface point P can be written (limited to a first-order approximation of the vertical gradient of the gravity anomaly, $\partial\Delta g/\partial h$)

$$\zeta(P) = \frac{R}{4\pi\gamma}\iint_\sigma S(\psi)\left[\Delta g + (h_P - h)\frac{\partial\Delta g}{\partial h}\right]\mathrm{d}\,\sigma \tag{1}$$

where R is the mean Earth radius, γ is normal gravity at normal height, σ is the unit sphere, $S(\psi)$ is Stokes's function with argument ψ as the geocentric angle, Δg is the surface gravity anomaly and h is the topographic elevation above the reference ellipsoid. Equation (1) is consistent with the first-order vertical gradient approximation of Bjerhammar (1962, 1963, 1969) formula for the quasi-geoid height

$$\zeta(P) = \frac{R}{4\pi\gamma}\iint_\sigma S(r_P,\psi)\left[\Delta g - h\frac{\partial\Delta g}{\partial h}\right]\mathrm{d}\,\sigma \tag{2}$$

where $S(r_P,\psi)$ is the extended Stokes function (Heiskanen and Moritz 1967, Sect. 6-4). Either of Eqs. (1) or (2) is taken to be the starting point for the RCR technique. However, as the gravity anomaly gradient is not known, it is common to approximate it by the expression (see e.g, Moritz 1968, 1980)

$$\begin{aligned}\frac{\partial\Delta g}{\partial h} &\approx \frac{R^2}{2\pi}\iint_\sigma \frac{\Delta g - \Delta g_P}{l_{PQ}^3}\,\mathrm{d}\sigma - \frac{2\Delta g}{R}\\ &\approx \frac{\mu R^2}{2\pi}\iint_\sigma \frac{h - h_P}{l_{PQ}^3}\,\mathrm{d}\,\sigma\end{aligned} \tag{3}$$

where l_{PQ} is the slope distance between point P at the Earth's surface and point Q at the sphere of radius R below the integration point and $\mu = G\rho$, i.e. gravitational constant times the (constant) topographic density. [Note that the first part of Eq. (3) will be exact if the gravity anomalies are located on a sphere.] In the last step of the approximation, the term $-2\Delta g/R$ has been neglected, and the gravity anomaly has been approximated by a linear correlation with topography, i.e.

$$\Delta g = a + 2\pi\mu h \tag{4}$$

where a is a constant. After a few more manipulations, the final, practical formula for the height anomaly is obtained as (Moritz 1980, Ch. 48)

$$\zeta(P) \approx \frac{R}{4\pi\gamma}\iint_\sigma S(\psi)\left(\Delta g + C_Q\right)\mathrm{d}\,\sigma - \frac{\pi\mu h_P^2}{\gamma} \tag{5a}$$

where C_Q is the well-known terrain correction given by

$$C_Q = \frac{\mu R^2}{2}\iint_\sigma \frac{(h_A - h_Q)^2}{l_{QA}^3}\,\mathrm{d}\,\sigma_A \tag{5b}$$

Finally, the geoid height (N) is obtained by adding the small correction (Heiskanen and Moritz 1967, p. 327)

$$N - \zeta \approx \frac{\Delta g_B}{\gamma}h \tag{6}$$

where

$$\Delta g_B = \Delta g - 2\pi\mu h \tag{7}$$

is the simple, planar Bouguer anomaly.

Frequently, the geoid height is directly derived using the classical Helmert method of condensation. For example, Sideris [1994a, Eqs. (1.11), (1.22), (1.23), 1994b] derived the approximate formula

$$N \approx \frac{R}{4\pi\gamma}\iint_\sigma S(\psi)\left[\Delta g + C_Q\right]\mathrm{d}\sigma_Q + \delta N_I \tag{8a}$$

where

$$\delta N_I \approx -\frac{\pi\mu}{\gamma}h^2 - \frac{\mu R^2}{6\gamma}\iint_\sigma \frac{h_A^3 - h_Q^3}{l_{AQ}^3}\,\mathrm{d}\sigma_A \tag{8b}$$

is the primary indirect topographic effect. Usually only the first term of δN_I is used in practice, causing an approximation error of upto 1 metre (Martinec and Vaníček 1994a).

In Sideris (1990, 1994a), the above Eqs. (8a) and (8b) were obtained by the approximations of Eq. (3) and the fact that the sum of the direct topographical effect and the downward continuation effect equals the terrain effect C_Q.

In the practical use of the RCR technique, the terrain correction [Eq. (5b)] is usually evaluated in planar approximation

$$c_Q = \frac{\mu R^2}{2} \iint_\sigma \frac{\left(h_A - h_Q\right)^2}{l_0^3} \, \mathrm{d}\sigma_A \tag{9a}$$

where $l_0 = 2R\sin(\psi/2)$, implying a difference from C_Q of the order of

$$\Delta C_Q = C_Q - c_Q \approx -\frac{3\mu R^2}{4} \iint_\sigma \frac{\left(h_A - h_Q\right)^4}{l_0^5} \, \mathrm{d}\sigma_A \tag{9b}$$

which can be most significant in rough terrain. With all these approximations, the final geoid estimator becomes

$$\hat{N} = \frac{R}{4\pi\gamma} \iint_\sigma S(\psi)\left(\Delta g + c_Q\right) \, \mathrm{d}\sigma - \frac{\pi\mu h^2}{\gamma} \tag{10}$$

At this point it is instructive to compare the basic Eqs. (5a) plus (6) and Eqs. (8a) plus (8b) for the geoid height estimation. In fact, the two approaches agree, except for the approximate correction $\Delta g_B h/\gamma$ of the first method. However, already this discrepancy is of the order of 1–2 m in the highest mountains. The contribution from the neglected term $-2\Delta g h/R$ of Eq. (3) when inserted into Eq. (2) is of the same order. Martinec and Vaníček (1994a); Martinec and Vaníč ek (1994b) and Sjöberg (2000, 2001) have emphasised that the resulting formulae for the Helmert condensation approach [Eqs. (8a) and (8b)] include several significant approximations in the direct and indirect topographic effects that may lead to errors of the order of several decimetres. This conclusion is well confirmed by a statement of Moritz concerning the last term of Eq. (5a) (which is of the order of 5 cm for elevations within 1 km but reaches 4 m for Mt. Everest): 'It will be consistent with the present approximation—restriction to first-degree corrections to Stokes' [sic] and Vening Meinesz' [sic] formulas and assumption of linear dependence of Δg on h—to neglect this small term' (1980, p. 418).

Interestingly, by assuming the Lipschitz condition that the slope of the terrain is within 45 degrees (i.e. the method fails in rough topography), and applying the refined surface Bouguer anomaly

$$\Delta g^{rB} = \Delta g - 2\pi\mu H - \tilde{C}_Q \tag{11a}$$

where

$$\tilde{C}_Q = \frac{\mu}{2} \iint_\sigma \frac{\left(h - h_Q\right)^2}{l_Q^3} \, \mathrm{d}\sigma \tag{11b}$$

and $l_Q = 2r_Q\sin(\psi/2) \approx l_0$ and using planar approximations [in their Eqs. (8) and (26)], Jekeli and Serpas (2003) showed that the geoid height can be determined by

$$N \approx \frac{R}{4\pi\gamma} \iint_\sigma S(\psi)\left[\Delta g + \tilde{C}_Q + \delta\Delta g_{dwc}^{rB}\right] \mathrm{d}\sigma + \delta N_I \tag{12a}$$

where

$$\delta\Delta g_{dwc}^{rB} = (\Delta g^{rB})^* - \Delta g^{rB} \tag{12b}$$

Here ( )* means that the quantity in the brackets is downward-continued to sea level, and δN_I is the indirect effect on the geoid for Helmert's method of condensation. The estimator in Eq. (12a) differs from the RCR estimator mainly by the term $\delta\Delta g_{dwc}^{rB}$. As $\tilde{C}_Q \approx C_Q$ and the refined Bouguer effect varies rather smoothly with elevation, one may assume that Eq. (12a) is a reasonable approximation of the geoid height. Nevertheless, it remains to prove whether it is good enough for the '1-cm geoid' required today. So far, it has not been proved that it is sufficient in high mountains, where the planar approximation may introduce decimetre errors, and the Lipschitz condition and the limitation of the formulas to first-order gravity anomaly gradients will not be adequate (see also Sect. 6.2).

Note that all the above has been derived for the full gravity field and the true topography, before taking out any global EGM reference field and topographic effect. This is what the RCR technique does, which will be discussed next.

3 The RCR technique

The concept of geoid determination by the RCR technique implies that both topography and low-degree gravity signals are removed before computation and restored after Stokes's integration. In addition, Stokes's integral is truncated to a limited region σ_0 around the computation point. In the approximations that follow, we will assume that σ_0 is a spherical cap of geocentric angle ψ_0 (although in the practical evaluations of the RCR technique the integration region is more often taken as a spherical rectangle). There are a number of methods called the RCR technique, but the two most well-known versions are probably the methods of residual terrain modelling (RTM) and Helmert's condensation, both of which will be breifly outlined below.

3.1 The RTM method

The RTM method (see e.g. Forsberg 1984, 1993, 1994; Forsberg and Tscherning 1997, Omang and Forsberg 2000) is based on the determination of the quasi-geoid height by analytical continuation [Eq. (2)], and the geoid height is obtained by adding the correction of Eq. (6). Formally, the RCR model for the height anomaly becomes

$$\zeta_{\mathrm{RCR}} = \frac{R}{4\pi\gamma} \iint_{\sigma_0} S(\psi) \times \left[\Delta g^M - \Delta g_{\mathrm{RTM}} + (h_P - h)\frac{\partial(\Delta g - \Delta g_{\mathrm{RTM}})}{\partial h}\right] \mathrm{d}\sigma + \zeta_M + \zeta_{\mathrm{RTM}} \tag{13a}$$

where

$$\Delta g^M = \Delta g - \Delta g_M = \Delta g - \sum_{n=2}^{M} \Delta g_n^{\mathrm{EGM}} \tag{13b}$$

is the residual gravity anomaly (assumed to give a signal of degrees higher than M) after subtracting the gravity anomaly Δg_M generated by an EGM, complete to degree and order M, and ζ_M is the corresponding low-degree EGM-generated quasi-geoid height. In practical evaluation, the gradient term under Stokes's integral is neglected, and

$$\Delta g_{\mathrm{RTM}} = 2\pi\mu(h - h_M) - c_Q \tag{14}$$

where h_M is a low-degree spherical harmonic representation of the topography generated by a global DTM and c_Q is the terrain correction of Eq. (9a) (applied in the planar approximation). Moreover, for ζ_{RTM}, the planar approximation

$$\zeta_{\mathrm{RTM}} = \frac{\mu}{\gamma} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(h - h_M)}{D} \, \mathrm{d}x \, \mathrm{d}y \tag{15}$$

is used, where D is the planar distance. Hence, the RCR estimator becomes

$$\hat{\zeta}_{\mathrm{RCR}} = \frac{R}{4\pi\gamma} \iint_{\sigma_0} S(\psi) \left(\Delta g^M + c_Q - 2\pi\mu(h - h_M)\right) \mathrm{d}\sigma + \zeta_M + \zeta_{\mathrm{RTM}} \tag{16}$$

and Stokes's integral is usually also evaluated in a planar approximation.

If we disregard the truncation error (to be considered in Sect. 5), and we assume that the RCR procedure of the EGM is exactly zero (see also next section), Eq. (16) should be consistent with Eq. (1). The possible error introduced by the estimator $\hat{\zeta}_{\mathrm{RCR}}$ with respect to Eq. (1) thus becomes

$$\varepsilon_{\hat{\zeta}} = \frac{R}{4\pi\gamma} \iint_{\sigma_0} S(\psi)[c_Q - g_1 - 2\pi\mu(h - h_M)] \, \mathrm{d}\sigma + \zeta_{\mathrm{RTM}} \tag{17}$$

where g_1 is the gravity anomaly gradient term of Eq. (1). From the spherical approximation of $\zeta_{\mathrm{RTM}}$and the approximation $S(\psi) \approx 2R/D$ (which only holds for small distances D), we easily obtain

$$\zeta_{\mathrm{RTM}} = \frac{\mu R^2}{\gamma} \iint_{\sigma_0} \frac{h - h_M}{D} \, \mathrm{d}\sigma \approx \frac{\mu R}{2\gamma} \iint_{\sigma_0} S(\psi)\,(h - h_M) \, \mathrm{d}\sigma \tag{18}$$

which thus exactly compensates the contribution $-2\pi\mu(h - h_M)$ from the RTM under Stokes's integral of Eq. (17). Numerical computations of the corresponding spherical approximations with cap sizes of 10 and 20 degrees [cf. Eq. (19a)] showed that this error is still within 1 cm. Hence, only the contribution from the gravity difference $c_Q - g_1$ remains, and Moritz (1980, Sect. 48) showed that Stokes's integral operating on this term yields approximately $\pi\mu h^2/\gamma$. This contribution reaches 1 cm for a computation point located at 430 m above sea level; for the height of Mt. Everest, it reaches 4.5 m. However, in Eq. (17) the integration area is limited to the spherical cap σ_0, which thus leads to the modified result

$$\varepsilon_{\hat{\zeta}} \approx \frac{\pi\mu}{\gamma} \sum_{n=2}^{\infty} \left(1 - \frac{n-1}{2} Q_n(\psi_0)\right) h_n^2 \tag{19a}$$

where h_n^2 is the nth Laplace spherical harmonic of h^2 (see e.g. Heiskanen and Moritz 1967, Sect. 1–13)

$$h_n^2 = \frac{2n+1}{4\pi} \iint_{\sigma} h^2 P_n(\cos\psi) \, \mathrm{d}\sigma \tag{19b}$$

and $Q_n(\psi_0)$ are Molodensky's truncation coefficients (Molodensky et al. 1962). A numerical application of Eq. (19a) for a cap size of 10 degrees and the upper limit of spherical harmonic series of the topography squared set to 1800 yielded a minimum, average and maximum global geoid error of –0.49, 0.01 and 2.52 m, respectively. The corresponding figures for a cap size of 5 degrees were –0.72, 0.01 and 2.45 m. This error is thus most significant in mountainous regions.

3.2 Helmert's second condensation approach

By removing and restoring an EGM and a DTM in Eq. (8a), we obtain (with obvious notations) the geoid estimator by Helmert condensation

$$N_{\mathrm{RCR}} = \frac{R}{4\pi\gamma} \iint_{\sigma_0} S(\psi)(\Delta g^M + C_Q - \Delta g_{\mathrm{DTM}}) \, \mathrm{d}\sigma + N_M + N_{\mathrm{DTM}} + \delta N_I \tag{20}$$

where C_Q and δN_I are determined by a DTM.

Disregarding the truncation bias of the integration area to σ_0 (to be considered in Sect. 5), we obtain the following error estimate by comparison with the more exact formula of Eq. (12a):

$$\varepsilon_{N_{\mathrm{RCR}}} \approx \frac{R}{4\pi\gamma} \iint_{\sigma_0} S(\psi)(C_Q - \tilde{C}_Q - \Delta g_M - \Delta g_{\mathrm{DTM}} - \delta\Delta g_{\mathrm{dwc}}^{rB}) \, \mathrm{d}\sigma + N_M + N_{\mathrm{DTM}} \tag{21}$$

Assuming that the total effect of the RCR procedure of the EGM vanishes and that $C_Q \approx \tilde{C}_Q$, the geoid error estimate becomes

$$\varepsilon_{N_{\mathrm{RCR}}} \approx -\frac{R}{4\pi\gamma} \iint_{\sigma_0} S(\psi)(\Delta g^{rB})_{\mathrm{dwc}} \, \mathrm{d}\sigma \tag{22}$$

which is likely to be a significant error at the several-centimetre level in mountainous regions. So far there are also several other significant topographic effects missing, which will be considered in Sect. 6.

3.3 General considerations

Usually the RCR technique employs a fast Fourier technique (FFT; see e.g. Forsberg 1985; Schwarz et al. 1990; Strang van Hees 1990; Forsberg and Sideris 1993 ; Haagmans et al. 1993; Sideris 1994a,b; Tziavos 1996) to speed up the computations. Importantly, the one-dimensional (1-D) FFT technique (Haagmans et al. 1993) can be used without introducing approximations to Stokes's integral, but the frequently used planar approximations inevitably introduce biased results.

There is no doubt that the use of a low-degree and low-order EGM can be very advantageous compared to the strict use of the original Stokes formula. This advantage stems from the assumption that the low-degree signal is better represented by the EGM, and Stokes's integral with a surface gravity anomaly is used only to compute the high degree signal. Whether this ambition is fulfilled in the RCR technique will be discussed in Sect. 4.

As shown in Sects. 2 and 3, the basic practical formulae for the RCR technique are too crude to allow a precise geoid determination of today in rough terrain and high mountains. Nevertheless, we may argue that the RCR technique could perhaps justify less accurate (topographic) corrections (see e.g. Tscherning 2001), because in this technique (1) the Stokes integration operates on a residual gravity anomaly rather than the total gravity anomaly signal, and (2) Stokes's integral is used in a limited region around the computation point. Remark (2) is certainly relevant for long-wavelength effects in the terrestrial gravity anomaly, but the short-wavelength effects, which dominate in the truncated Stokes integral, cannot be avoided in this way. Remark (1) is discussed in Sect. 4 with respect to applying a higher-order reference field.

However, at this point, we should admit that the use of the residual gravity anomaly is advantageous in smoothing and interpolating gravity anomaly data used in Stokes's integral, and, as discussed in Sect. 5, the removal of the RTM reduces the truncation error in Stokes's integral. Nevertheless, the RCR procedure of the RTM must be consistent, so as not to introduce additional biases to the estimation procedure. As shown in Sect. 3.1, this is not the case with the RTM method as it is usually applied today.

4 The higher-order reference field

The two main questions to be answered here are whether (1) the RCR technique of Eqs. (16) and (20), operating with residual gravity anomalies, can be used in Stokes's formula with less precise corrections than the original, but truncated, Stokes's formula, which uses the full gravity anomaly signal, and (2) whether Eqs. (16) and (20) efficiently exploit the EGM for the long-wavelength gravity field representation. To answer these questions we will only consider the use of a higher-order reference field, which by itself reduces the gravity anomaly under the integral by a considerable amount. To obtain the answers, we will focus our discussion on a comparison of the following two estimators of the geoid height (cf. Sjöberg 1986):

$$\tilde{N}_1 = \frac{R}{4\pi\gamma}\iint\limits_{\sigma_0} S(\psi)\,\Delta g^M\,\mathrm{d}\sigma + c\sum_{n=2}^{M}\frac{2}{n-1}\Delta g_n^{\mathrm{EGM}} \tag{23}$$

and

$$\tilde{N}_2 = \frac{R}{4\pi\gamma}\iint\limits_{\sigma_0} S(\psi)\,\Delta g\,\mathrm{d}\sigma + c\sum_{n=2}^{M}Q_n\Delta g_n^{\mathrm{EGM}} \tag{24}$$

where $c = R/2\gamma$, $\Delta g_n^{\mathrm{EGM}}$ is the Laplace harmonic of degree n for the gravity anomaly determined from the EGM, and $Q_n = Q_n(\psi_0)$. The first estimator [Eq. (23)], originating with Vincent and Marsh (1974), is of the form with a higher-order reference field that uses the residual gravity anomaly Δg^M in Stokes's formula, while the second estimator [Eq. (24); Rapp and Rummel 1975] utilises the full surface gravity anomaly inside the integration cap and the EGM to compute the effect outside the cap. The estimator $\tilde{N}_1$ is thus apparently advantageous in comparison with to the estimator $\tilde{N}_2$, because since $\tilde{N}_1$ uses the residual gravity anomaly Δg^M, it may be expected to be less affected by any gravity anomaly error than $\tilde{N}_2$, which uses the full gravity anomaly signal Δg. However, this type of reasoning is completely wrong, as we will show next in two ways.

First, let us assume that the surface gravity anomaly is in error by $\varepsilon_{\Delta g}$. It is easy to see that this error propagates into the following errors of the geoid estimators [Eqs. (23) and (24)]:

$$\delta\tilde{N}_1 = \frac{R}{4\pi\gamma}\iint\limits_{\sigma_0} S(\psi)\,\varepsilon_{\Delta g}\,\mathrm{d}\sigma = \delta\tilde{N}_2 \tag{25}$$

i.e. the two estimators have the same error.

Second, by introducing the spectral form of the terrestrial gravity anomaly, $\Delta g = \sum \Delta g_n^T$, the two estimators can be written in their spectral forms as follows. From $\tilde{N}_1$ we obtain

$$\tilde{N}_1 = c\sum_{n=2}^{\infty}\left[\frac{2}{n-1} - Q_n\right]\Delta g_n^T + c\sum_{n=2}^{M}Q_n\Delta g_n^{\mathrm{EGM}} \tag{26}$$

which is nothing other than the spectral form of $\tilde{N}_2$, and we have thereby shown that the two estimators are actually identical. [Note that this identity between $\tilde{N}_1$ and $\tilde{N}_2$ holds for any error distribution of the gravity anomalies and potential coefficients, and it holds also for the modified Stokes function (see Appendix A).] Hence, the answer to our first question is in the negative, and there is therefore no particular advantage in using a higher-order reference field compared to a direct combination of Stokes's integral with an EGM.

On the other hand, there is no doubt that there could be a great advantage in combining Stokes's formula with a set of EGM potential coefficients, which is the case in both Eqs. (23) and (24). This combination is then intended to use the long-wavelength information of the EGM, and the short-wavelength features are to be estimated by Stokes's formula in a spherical cap or other (usually rectangular with FFT) area around the computation point. However, it also follows from Eq. (26) that the answer to our second question (2) is in the negative: the RCR method does not fully utilise the long-wavelength gravity field from the EGM: the EGM is only used for the contribution from the remote zone. Hence, the RCR method does not have the desired property of removing all low-degree terrestrial gravity signal and representing it by the EGM, it only reduces the remote-zone truncation error (in a non-optimum way) (see also Vaníček and Featherstone 1998; Featherstone 2003). With the recent dedicated gravimetric satellite missions, such as CHAMP, GRACE and the near-future GOCE missions, the low-degree gravity field (say to $M = 200$) will be accurately determined with satellite-only EGMs; however, as shown by Eq. (26), the RCR technique will fail to take full advantage of such high-precision data. [In contrast, the requested satellite-only low-degree combinations with terrestrial gravity anomaly can be achieved by the so-called modified Stokes formula for a special choice of modification parameters (Vaníček and Sjöberg 1991; Sjöberg 2003b).]

5 The truncation error

The estimator $\tilde{N}_1$ (or $\tilde{N}_2$) in Sect. 4 suffers from the truncation of the EGM to the maximum degree M and of the Stokes integration to a cap/rectangle σ_0. Considering only the truncation of Stokes's integral, we may define the true value for the geoid height by

$$N = \frac{c}{2\pi} \sum_{n=2}^{\infty} \frac{\Delta g_n}{n-1} \tag{27}$$

and the truncation error, using the spectral form of the estimator [Eq. (26)], can then be written

$$\delta\tilde{N}_1 = \tilde{N}_1 - N = -c \sum_{n=M+1}^{\infty} Q_n \Delta g_n \tag{28}$$

Equation (28) can be estimated in the global RMS sense by the formula

$$\delta\bar{N}_1 = c\sqrt{\sum_{n=M+1}^{\infty} Q_n^2 c_n} \tag{29}$$

where c_n are the gravity anomaly degree variances. Truncation errors for $M = 360$ and different sizes of the cap radius, which have been computed using Eq. (29) and the gravity anomaly degree variances of Tscherning and Rapp (1974), are presented in the second column of Table 1. This is the RMS truncation error of Eq. (20), i.e. with Helmert's second method of condensation.

Hence, to keep the truncation bias below 1 cm and using the original (unmodified) Stokes's function, the cap radius must exceed approximately 20 degrees. However, in practice, the size of the rectangular integration area frequently corresponds to a spherical cap radius of 5–10 degrees. According to Table 1, column 2, this yields a significant truncation bias, depending on position. In column 3 of Table 1, we show the RMS truncation error for the RTM method with degree variances taken from Forsberg (1986) and Tscherning and Forsberg (1987). It shows some reduction of the truncation error compared with Helmert's method, but the error is significant at the centimetre level for any cap size less than 10 degrees.

Table 1 Global RMS truncation errors for different methoods with $L = M = 360$ and cap sizes ψ_0 in degrees. Unit: cm

ψ_0	Helmert	RTM	Molodensky
2	3.2	2	0.0
5	2.3	1.4	0.0
10	1.6	1.0	0.0
20	0.9	0.6	0.0

However, by using a method of modifying Stokes's kernel (Molodensky et al. 1962; Sjöberg 1984), the goal of an insignificant truncation bias can be achieved for a much smaller cap size when utilising an EGM to degree and order $M = 360$. Let us illustrate this for Molodensky's method of modifying Stokes's formula. We start from

$$N_3 = \frac{R}{4\pi\gamma} \iint_{\sigma_0} S^L(\psi)\, \Delta g \, d\sigma + c \sum_{n=2}^{M} \left(Q_n^L + s_n\right) \Delta g_n^{\mathrm{EGM}} \tag{30}$$

with the modified Stokes function

$$S^L(\psi) = S(\psi) - \sum_{n=2}^{L} \frac{2n+1}{2} s_n P_n(\cos\psi) \tag{31}$$

where Q_n^L are Molodensky truncation coefficients using the kernel in Eq. (31) modified to any degree L. The parameters s_n are determined in such a way that the upper limit of the truncation error is minimised (Molodensky et al. 1962; Sjöberg 1984). This leads to an expression for the truncation error that is analogous to Eq. (28). The global RMS $\delta\bar{N}_3$ is given by a formula similar to Eq. (29). Again, utilising the Tscherning and Rapp (1974) degree variances and $L = M = 360$, we arrive at near-zero truncation errors (Table 1, column 4). It is thus clear that a significantly smaller cap size can be used in this case. However, in reality, our gravity observations and EGM geopotential coefficients contain errors, which affect both the choice of cap size and how Stokes's formula should be modified (see e.g. Sjöberg 1984, 1991).

6 The topographic effects

As discussed in Sect. 3.1, the RTM method is primarily biased by the term $\pi\mu h^2/\gamma$, which is easy to correct by applying a correction -$\varepsilon_{\hat{\zeta}}$ [Eq. (19)]. Similarly, the Helmert-type geoid estimator [Eq. (20)] can be improved by the correction $-\varepsilon_{\hat{N}_{\mathrm{RCR}}}$ [Eq. (22)]. Here we will discuss the additional topographic corrections needed in rough terrain.

6.1 The RTM method

The correction $-\varepsilon_{\hat{\zeta}}$ was derived by assuming a linear correlation between the gravity anomaly and elevation, an approximation that might be insufficient and generally not true (Martinec 1998). Moreover, the basic Eq. (1), based on the first-order Molodensky series, might be insufficient in such terrain, and the correction $N - \zeta$ of Eq. (6) may also be too crude (see e.g. Sjöberg 1995). Finally, the assumption of a constant topographic density will not necessarily suffice. All these effects must be considered for the goal of the '1-cm geoid'.

6.2 Helmert's second method of condensation

The Helmert condensation technique by the geoid estimator of Eq. (20) with the correction $-\varepsilon_{\hat{N}}$ of Eq. (22) is not adequate for 1-cm geoid estimation in rough terrain unless the following additional significant topographic effects are considered.

1. The terms of Eq. (20) representing the EGM must also be corrected for the direct topographic effect (Sjöberg 1996, 2001). (This is obvious if we consider the special case of these equations with $\sigma_0 = 0$, as the geoid height represented by an EGM will be biased when predicting the geoid below the topographic surface.)

2. Although Stokes's function is blind to harmonics of degrees zero and one, the geoid estimation needs topographic corrections for these harmonics (Sjöberg 2001). These corrections are of the order of 10 cm.

3. In high mountain areas, the limitation of the topographic corrections to first-order approximations will not be sufficient, and the exact surface integrals of direct and indirect effects presented in Martinec (1998) and Sjöberg (2000) should be useful.

4. The assumption of a constant topographic density is not sufficient in mountainous regions (Martinec 1998), but a correction for a laterally variable topographic density can be applied (also see Sjöberg 2004).

7 Concluding remarks

We have shown that the use of a higher-order reference field, as is the case for the RCR technique, does not make the geoid estimators more insensitive to various errors than does a combined solution of Stokes's formula. Admittedly, one exception is that the use of a residual gravity anomaly reduces the numerical errors committed by the Stokes integration. The use of a higher-order reference field without modifying Stokes's kernel does not take advantage of the high-quality low-degree signal of the EGM in an efficient way; low-degree gravity anomaly errors will deteriorate the solution. In fact, the RCR technique, as it has been applied to date, uses the EGM just to reduce the truncation error in a non-optimum way, and some of the potential advantages of the EGM are lost. In contrast to the RCR technique, it is only by the modification of Stokes's kernel (see e.g. Sjöberg 1991 and 2003a) that it is possible to match the errors of the terrestrial gravity data, the EGM and the truncation in an optimum way. Consequently, in order to cope with the present goals for precise regional geoid determination, the RCR technique must also employ the modified Stokes's kernel technique and all the refined topographic and atmospheric and other corrections to gravimetric geoid determination available today. Otherwise, a precise EGM that will result from the dedicated satellite gravity and gradiometry missions, in combination with the RCR technique, will most probably not solve the problems and achieve the desired 1-cm geoid model.

Acknowledgements Jonas Ågren contributed with some essential remarks to a previous version of the manuscript. Jonas Ågren and Artu Ellmann provided the numerical results for this paper. This support is cordially acknowledged.

Appendix A

Consider the two models for estimating geoid heights

$$\tilde{N}_1 = \frac{c}{2\pi} \iint_{\sigma_0} S^L(\psi)\, \Delta\tilde{g}^M \, d\sigma + c \sum_{n=2}^{M} \frac{2}{n-1} \Delta\tilde{g}_n^{\mathrm{EGM}} \tag{A1}$$

and

$$\tilde{N}_2 = \frac{c}{2\pi} \iint_{\sigma_0} S^L(\psi)\, \Delta\tilde{g} \, d\sigma + c \sum_{n=2}^{M} \left(Q_n^L + s_n^*\right) \Delta\tilde{g}_n^{\mathrm{EGM}} \tag{A2}$$

where

$$\Delta\tilde{g}^M = \Delta\tilde{g} - \Delta\tilde{g}^{\mathrm{EGM}} \tag{A3}$$

$$S^L(\psi) = S(\psi) - \sum_{n=2}^{L} \frac{2n+1}{2} s_n P_n(\cos\psi) \tag{A4}$$

and

$$s_n^* = \begin{cases} s_n & \text{if } 2 \le n \le L \\ 0 & \text{otherwise} \end{cases} \tag{A5}$$

where L and M are arbitrary upper limits of summation of modification of Stokes's kernel and the EGM, respectively, and $\tilde{x}$ denotes an erroneous estimate of any quantity x.

Proposition $\tilde{N}_1 = \tilde{N}_2$.

Proof By considering Eq. (A3), $\tilde{N}_1$ can be rewritten

$$\begin{aligned} \tilde{N}_1 &= \frac{c}{2\pi} \iint_{\sigma_0} S^L(\psi)\, \Delta\tilde{g} \, d\sigma - \frac{c}{2\pi} \iint_{\sigma_0} S^L(\psi)\, \Delta\tilde{g}^{\mathrm{EGM}} \, d\sigma \\ &\quad + c \sum_{n=2}^{M} \frac{2}{n-1} \Delta\tilde{g}_n^{\mathrm{EGM}} \\ &= \frac{c}{2\pi} \iint_{\sigma_0} S^L(\psi)\, \Delta\tilde{g} \, d\sigma + c \sum_{n=2}^{M} \left(Q_n^L + s_n^*\right) \Delta\tilde{g}_n^{\mathrm{EGM}} \end{aligned} \tag{A6}$$

because

$$\frac{c}{2\pi}\iint\limits_{\sigma_0} S^L(\psi)\,\Delta\tilde{g}^{\mathrm{EGM}}\,\mathrm{d}\sigma$$
$$= c\sum_{n=2}^{M}\left(\frac{2}{n-1} - Q_n^L - s_n^*\right)\Delta\tilde{g}_n^{\mathrm{EGM}} \tag{A7}$$

Comparing the last term of Eq. (A6) and Eq. (A2) proves that $\tilde{N}_1$ and $\tilde{N}_2$ are equal.

References

Bjerhammar A (1962) Gravity reduction to a spherical surface. Rep, Division of Geodesy, Royal Institute of Technology, Stockholm

Bjerhammar A (1963) A new theory of gravimetric geodesy. Rep, Division of Geodesy, Royal Institute of Technology, Stockholm

Bjerhammar A (1969) On the boundary value problem of physical geodesy. Tellus 21(4):451–516

Denker H, Behrand D, Torge W (1996) The European gravimetric quasigeoid EGG95. In: Sansó F (ed) New geoids of the world. Int Geoid Service Bulletin 4, Milano, pp 3–12

Featherstone WE (2003) Tests of two forms of Stokes's integral using a synthetic gravity field based on spherical harmonics. In: Grafarend EW, Krumm FW, Schwarze VS (eds) Geodesy—the challenge of the Third Millennium. Springer, Berlin Heidelberg New York, pp 163–171

Forsberg R (1984) A study of terrain reductions, density anomalies and geophysical inversion methods in gravity field modelling. Rep 355, Department of Geodetic Science, The Ohio State University, Columbus

Forsberg R (1985) Gravity field terrain effect computations by FFT. Bull Géod 59:342–360

Forsberg R (1986) Spectral properties of the gravity field in the Nordic countries. Proc Int Symp on the Definition of the Geoid, Florence, 26–30 May, pp 295–324

Forsberg R (1990) A new high-resolution geoid of the Nordic area. In: Rapp RH, Sansó F (eds) Determination of the geoid. Springer, Berlin Heidelberg New York, pp 241–250

Forsberg R (1993) Modelling the fine-structure of the geoid: methods, data requirements and some results. Surv Geophys 14:403–418

Forsberg R (1994) Terrain effects in geoid computation. In: Lecture notes, International School for the Determination and Use of the Geoid. Int Geoid Service, DIIAR—Politecnico di Milano, pp 101–134

Forsberg R (2001) Development of a Nordic cm-geoid—with basics of geoid determination. In: Harsson BG (ed) Lecture notes for autumn school organised by the Nordic Geodetic Commission, Fevik, 28 August—2 September 2000. Statens kartverk, Hønefoss, Norway

Forsberg R, Sideris MG (1993) Geoid computation by the multi-band spherical FFT approach. Manuscr Geod 18:82–90

Forsberg R, Tscherning CC (1997) Topographic effects in gravity field modelling for BVP. In Sansó F, Rummel R (eds) Geodetic boundary value problems in view of the one centimeter geoid. Lecture Notes in Earth Sciences 65. Springer, Berlin Heidelberg New York, pp 121–160

Haagmans R, de Min E, Van Gelderen M (1993) Fast evaluation of convolution integrals on the sphere using 1D FFT and a comparison with existing methods for Stokes' integral. Manuscr Geod 18:227–241

Heiskanen WA, Moritz H (1967) Physical geodesy. WH Freeman, San Francisco

Jekeli C, Serpas JG (2003) Review and numerical assessment of the direct topographical reduction in geoid determination. J Geod 77:226–239

Martinec Z (1998) Boundary-value problems for gravimetric determination of a precise geoid. Lecture Notes in Earth Sciences 73. Springer, Berlin Heidelberg New York

Martinec Z, Vaníček P (1994a) The indirect effect of topography in the Stokes–Helmert technique for a spherical approximation of the geoid. Manuscr Geod 19:213–219

Martinec Z, Vaníček P (1994b) Direct topographical effect of Helmert's condensation for a spherical geoid. Manuscr Geod 19:257–268

Milbert DG (1996) Improvement of a high resolution geoid model in the United States by GPS height on NAVD88 benchmarks. In Sansò F (ed) New geoids of the world. Int Geoid Service Bulletin 4, Milano, pp 13–36

Molodensky MS, Eremeev VF, Yurkina MI (1962) Methods for study of the external gravity field and figure of the earth. Transl from Russian (1960). Israel Program for Scientific Translations, Jerusalem

Moritz H (1966) Linear solutions to the geodetic boundary-value problems. Rep 79, Department of Geodetic Science, Ohio State University, Columbus

Moritz H (1968) On the use of the terrain correction in solving Molodensky's problem. Rep 108, Department of Geodetic Science, The Ohio State University Columbus

Moritz H (1980) Advanced physical geodesy. Wichmann, Karlsruhe

Omang OCD, Forsberg R (2003) How to handle the topography in geoid determination: three examples. J Geod 74:458–466

Rapp RH, Rummel R (1975) Methods for the computation of detailed geoids and their accuracy. Rep 233, Department of Geodetic Science, The Ohio State University, Columbus

Sansó F (1994) (ed) Lecture notes, International School for the Determination and Use of the Geoid. International Geoid Service, DIIAR—Politechnico di Milano

Sansó F, Rummel R (eds) (1997) Geodetic boundary value problems in view of the one centimetre geoid. Lecture Notes in Earth Sciences 65. Springer, Berlin Heidelberg New York

Schwarz K-P, Sideris MG, Forsberg R (1990) Use of FFT methods in physical geodesy. Geophys J Int 100:485–514

Sideris MG (1990) Rigorous gravimetric terrain modelling using Molodensky's operator. Manuscr Geod 15:97–106

Sideris MG (1994a) Geoid determination by FFT techniques. In: Sanso F (ed) Lecture notes, International Geoid School for the determination and use of the geoid. International Geoid Service, DIIAR—Politechnico di Milano

Sideris MG (1994b) Regional geoid determination. In: Vaníček P, Christou NT (eds) The geoid and its geophysical interpretation. CRC Press, Boca Raton, FL, pp 77–94

Sjöberg LE (1984) Least squares modification of Stokes' and Vening Meinesz' formulas by accounting for truncation and potential coefficient errors. Manuscr Geod:9:209–229

Sjöberg LE (1986) Comparison of some methods of modifying Stokes' formula. Proc 10th General Meeting of the NKG, pp 268–278. Also in Boll Geod Sci Aff 46(2):229–248

Sjöberg LE (1991) Refined least squares modification of Stokes' formula. Manuscr Geod 16:367–375

Sjöberg LE (1995) On the quasigeoid to geoid separation. Manuscr Geod 20:182–192

Sjöberg LE (1996) The terrain effect in geoid computation from satellite derived geopotential models. Boll Geod Sci Aff 55:385–392

Sjöberg LE (2000) On the topographic effects by the Stokes–Helmert method of geoid and quasi-geoid determinations. J Geod 74(2):255–268

Sjöberg LE (2001) Topographic and atmospheric corrections of gravimetric geoid determination with special emphasis on the effects of degree zero and one. J Geod 75:283–290

Sjöberg LE (2003a) A general model of modifying Stokes' formula and its least squares solution. J Geod 77:459–464

Sjöberg LE (2003b) Improving modified Stokes' formula by GOCE data. Boll Geod Sci Aff 61(3):215–225

Sjöberg LE (2004) The effect on the geoid of lateral density variations. J Geod 78:34–39

Sjöberg LE, Å gren J (2002) Some problems in the theory used for the NKG geoid model. In: Poutanen M, Suurmäki H (eds) Proc 14th General Meeting of the Nordic Geodetic Commission. Finnish Geodetic Institute, Masala, pp 121–130

Strang van Hees G (1990) Stokes formula using fast Fourier techniques. Manuscr Geod 15:235–239

Torge W (2001) Geodesy, 3rd edn. De Gruyter, Berlin
Tscherning CC (2001) Communication. The Hotine–Marussi Symposium on Mathemathical Geodesy
Tscherning CC, Forsberg R (1986) Geoid determination in the Nordic countries—a status report. In: Kakkuri J (ed) Proc 10th General Meeting of the NKG, Helsinki, 29 September—3 October. Finnish Geodetic Institute, Masala
Tscherning CC, Forsberg R (1987) Geoid determination in the Nordic countries from gravity and height data. Boll Geod Sci Aff 46:21–43
Tscherning CC, Rapp RH (1974) Closed covariance expressions for gravity anomalies, geoid undulations, and deflections of the vertical implied by anomaly degree variance models. Rep 355, Department of Geodetic Science, The Ohio State University, Columbus
Tziavos IN (1996) Comparisons of spectral techniques for geoid computations over large regions. J Geod 70(6):357–373
Vaníček P, Featherstone WE (1998) Performance of three types of Stokes's kernels in the combined solution for the geoid. J Geod 72(12):684–697
Vaníček P, Sjöberg LE (1991): Reformulation of Stokes's theory for higher than second-degree reference field and modification of integration kernels. J Geophy Res 96(B4):6529–6539
Vincent S, Marsh J (1974) Gravimetric global geoid. In: Veis G (ed) Proc Int Symp On the Use of Artificial Satellites for Geodesy and Geodynamics, National Technical University, Athens