Estimating divergence from irregularly spaced observations: A comparison of three techniques

Jacqueline A. Dubois

REU and School of Meteorology, University of Oklahoma, Norman, Oklahoma

Phillip L. Spencer*

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

*Also affiliated with the NOAA/OAR National Severe Storms Laboratory, Norman, OK

Research Experiences for Undergraduates Final Project

31 July 2004

Corresponding author address: Jacqueline A. Dubois, 408 Settlers Drive, Lawrence, KS 66049
Email: jackiedubois@ou.edu

Abstract

Three methods for estimating gridded fields of divergence from irregularly spaced wind observations are evaluated by sampling analytic fields of cyclones and anticyclones of varying wavelengths using a surface network. For the *finite differencing* method, divergence is computed from the objectively analyzed horizontal wind field. For the *triangle method*, which requires a triangular tessellation of the station network and assumes that the wind varies linearly within each triangle, divergence estimates are obtained directly from the wind observations and are assumed valid at triangle centroids. These irregularly spaced centroid divergence estimates then are analyzed to a grid. For the *pentagon method*, which requires a pentagonal tessellation of the station network and assumes that the wind varies quadratically within each pentagon, divergence estimates are obtained directly from the wind observations and are valid at the station lying within the interior of each pentagon. These irregularly spaced divergence estimates then are analyzed to a grid.

Results support previous studies showing that the triangle method provides analyses with lower root-mean-square errors that those of the finite differencing method. We find that for all wavelengths considered, the triangle method provides better analyses than the pentagon method, as well, despite the more restrictive assumption by the triangle method regarding the wind field. However, for most wavelengths considered, we find that the divergence estimates at the interiors of pentagons are superior to those at triangle centroids.

1. Introduction

Obtaining accurate estimates of spatial derivatives from observed data is important in meteorology. For example, divergence, comprised of a combination of spatial derivatives of the wind field, appears in numerous diagnostic equations such as those involving mass continuity, vorticity tendency, and moisture convergence. Thus, the accuracy of the diagnosis depends on the accuracy of the divergence estimates. Because many techniques are available for estimating gridded fields of divergence from irregularly spaced wind observations, it is important to understand the strengths and weaknesses of each technique and the conditions under which each method may be preferred.

The purpose of this study is to compare three techniques for computing gridded fields of divergence: the finite differencing method, the triangle method, and the pentagon method. Each of the three methods involves two basic procedures: derivative estimation and the application of an analysis scheme. An important difference between the three techniques is that these two basic procedures are not performed in the same order. Specifically, for the finite differencing method, an analysis scheme is applied to the wind observations and then a finite differencing scheme is used to estimate divergence. In contrast, for the triangle and pentagon methods, divergence estimates are obtained *directly* from the wind observations and then an analysis scheme is used to map the estimates onto a grid.

Previous studies have shown that the triangle method for obtaining gridded divergence estimates generally is superior to the finite differencing method (e.g., Schaefer and Doswell 1979; Doswell and Caracena 1988; Spencer and Doswell 2001,

hereafter SD01). Much of this study follows the same approach as SD01, whereby a comparison between the analyses and an analytically-prescribed divergence field quantifies the effectiveness of each of the methods. In contrast to SD01, our work 1) incorporates a real observation network and 2) evaluates the pentagon method, comparing its analyses to those of the finite differencing and triangle methods.

Section 2 describes how analytic observations are created. Section 3 gives a description of each of the three methods considered. The results of the evaluations are presented in section 4. Section 5 contains a brief summary.

2. Creating analytic observations

Observations are created by sampling analytic functions at the stations indicated in Fig. 1. The advantage of using analytic wind fields is that the true divergence values are known everywhere in the region, making it easy to compare the relative merits of the three methods. The stations in Fig. 1 represent a set of declustered surface observation locations whose average separation (Δ) is approximately 125 km. The horizontal components of the analytic wind field (u_a,v_a) are represented by the following equations:

$$u_a(x,y) = 10\cos\left(\frac{2\pi}{L}x - \phi_x\right)\sin\left(\frac{2\pi}{L}y - \phi_y\right) \qquad (1a)$$

$$v_a(x,y) = 10\sin\left(\frac{2\pi}{L}x - \phi_x\right)\cos\left(\frac{2\pi}{L}y - \phi_y\right), \qquad (1b)$$

where L is the wavelength. From (1), the analytic divergence (δ_a) is easily derived as:

$$\delta_a(x,y) = -20\left(\frac{2\pi}{L}\right)\sin\left(\frac{2\pi}{L}x - \phi_x\right)\sin\left(\frac{2\pi}{L}y - \phi_y\right). \qquad (2)$$

This represents a checkerboard pattern of cyclones and anticyclones.

3. Methods for computing gridded fields of divergence

a. Finite differencing method

For the finite differencing method, wind observations are analyzed to a 34x30 grid using a 3-pass Barnes objective analysis scheme (hereafter, BOA3; Barnes 1964, 1973; Achtemeier 1987, 1989). The BOA3 weighting function is given by

$$w_k = e^{\frac{-R_k^2}{(c_i \Delta)^2}}, \qquad (3)$$

where w_k is the weight of the kth observation, R_k is the distance between an observation and a gridpoint, and c_i is the smoothing parameter used during the ith analysis pass. The nominal grid spacing is 40 km.

Following the suggestion of Achtemeier (1989), we use a relatively large value of the smoothing parameter for the first pass ($c_1 = 1.75$). The shape parameters for the two correction passes are chosen such that $c_2 = c_3 \equiv c_{2,3}$. In order to determine the value of $c_{2,3}$ that provides the best analysis, as measured by root mean square error and correlation coefficient (section 4a), $c_{2,3}$ is varied from 0.5 to 2.5 in increments of 0.1. No "cutoff radius" is used in our analysis scheme.

After BOA3 is applied, a fourth-order centered finite differencing scheme is used to estimate the divergence of the gridded wind field.

b. Triangle method

The triangle method requires a triangular tessellation of the station network. To accomplish this, the Delauney triangulation scheme is used to create a set of non-overlapping triangles (Ripley 1981). Four nearby stations comprise a quadrilateral, of

which competing triangular tessellations exist. The Delauney triangulation scheme selects the set of triangles with the largest minimum angle from the two competing tessellations. Also, the scheme requires that the nearest observation from any point within a given triangle be one of the vertices of that triangle. The triangular tessellation of the station network shown in Fig. 1 is presented in Fig. 2.

The triangle method assumes a linear variation of the wind field within each triangle. Therefore, the horizontal wind components at the *i*th station comprising a triangle (u_i,v_i) can be represented by a linear Taylor series:

$$u_i(x,y) = u_c + \left.\frac{\partial u}{\partial x}\right|_c \Delta x_i + \left.\frac{\partial u}{\partial y}\right|_c \Delta y_i \qquad (4a)$$

$$v_i(x,y) = v_c + \left.\frac{\partial v}{\partial x}\right|_c \Delta x_i + \left.\frac{\partial v}{\partial y}\right|_c \Delta y_i \quad , \qquad (4b)$$

where (u_c, v_c) represent the horizontal wind components at the triangle centroid located at (x_c, y_c), $\Delta x_i=(x_i-x_c)$, and $\Delta y_i=(y_i-y_c)$. For each triangle, the system of six equations and six unknowns is solved for the unknown gradients, from which divergence is computed. The irregularly spaced divergence estimates at the triangle centroids then are analyzed to the 34x30 grid using BOA3. It is important to note that divergence values from the triangle method are estimated *directly* from the wind observations and then analyzed to the grid.

c. Pentagon method

The pentagon method requires a pentagonal tessellation of the station network. Since no automated routine was available, this task was performed by hand such that each observing station (except those along the border of the domain) is enclosed within

five nearby stations forming a pentagon. The general guidelines for forming each pentagon were to 1) minimize the area of the pentagon and 2) place the interior station as near as possible to the center of the pentagon. A few such pentagons are shown in Fig. 3. Following Chien and Smith (1973), the horizontal wind components at the *i*th station comprising a pentagon (u_i,v_i) can be represented by a second-order Taylor series:

$$u_i(x,y) = u_0 + \left.\frac{\partial u}{\partial x}\right|_0 \Delta x_i + \left.\frac{\partial u}{\partial y}\right|_0 \Delta y_i + \frac{1}{2!}\left.\frac{\partial^2 u}{\partial x^2}\right|_0 \Delta x_i^{\,2} + \left.\frac{\partial^2 u}{\partial x \partial y}\right|_0 \Delta x_i \Delta y_i + \frac{1}{2!}\left.\frac{\partial^2 u}{\partial y^2}\right|_0 \Delta y_i^{\,2} \qquad (5a)$$

$$v_i(x,y) = v_0 + \left.\frac{\partial v}{\partial x}\right|_0 \Delta x_i + \left.\frac{\partial v}{\partial y}\right|_0 \Delta y_i + \frac{1}{2!}\left.\frac{\partial^2 v}{\partial x^2}\right|_0 \Delta x_i^{\,2} + \left.\frac{\partial^2 v}{\partial x \partial y}\right|_0 \Delta x_i \Delta y_i + \frac{1}{2!}\left.\frac{\partial^2 v}{\partial y^2}\right|_0 \Delta y_i^{\,2}, \qquad (5b)$$

where (u_0, v_0) are the horizontal wind components at the interior station located at (x_0, y_0), $\Delta x_i=(x_i-x_0)$, and $\Delta y_i=(y_i-y_0)$. The system of ten equations and ten unknowns is solved to yield an estimate of the divergence at the interior station of each pentagon. The irregularly spaced divergence estimates then are analyzed to the grid using BOA3. We emphasize that, as in the triangle method, the divergence values are estimated *directly* from the observations and then analyzed to the grid.

4. Comparison of the three methods

The root-mean-square error (rmse) and Pearson correlation coefficient are used to quantify the results of the three methods. Each method is applied to four variations of the analytic wind field (L=5Δ, L=10Δ, L=15Δ, and L=20Δ) using various values of $c_{2,3}$ as described in section 3a. The correlation coefficient provides a better measure of the correctness of the pattern of an analysis than does the rmse, whereas the rmse provides a better measure of correctness of the amplitude of an analysis. These error statistics are calculated at the triangle centroids and interior stations of pentagons (i.e., before BOA3

scheme is applied) as well as from the gridded analyses (i.e., after the BOA3 scheme is applied). We define the "best" method at each of the wavelengths as that which produces the lowest rmse and/or the highest correlation coefficient.

a. Error estimation

The root-mean-square error (rmse) computed from an analysis is defined as follows:

$$rmse = \sqrt{\frac{\sum_{i,j} (\delta_g - \delta_a)^2}{N_g}}, \tag{6}$$

where N_g represents the number of grid points in the verification domain, δ_g represents a divergence analysis, and δ_a represents the analytic divergence at gridpoints. In order to avoid contamination of the rmse by boundary errors, the verification domain is limited to the innermost one-half of the analysis domain. A similar equation is used to compute rmses at triangle centroids and at the interior stations of pentagons.

The correlation coefficient computed from an analysis is defined as follows:

$$\text{correlation coefficient} = \frac{\sum_{i,j} (\delta_g - \overline{\delta_g})(\delta_a - \overline{\delta_a})}{\left[\sum_{i,j} (\delta_g - \overline{\delta_g})^2 \sum_{i,j} (\delta_a - \overline{\delta_a})^2 \right]^{1/2}}, \tag{7}$$

where $\overline{\delta_g}$ represents the average value of the analysis and $\overline{\delta_a}$ represents the average value of the analytic field (Wilks 1995). This statistic is calculated over the same domain as the rmse.

b. Results

Figures 4 and 5 illustrate the differences in the three methods for L=10Δ and L=20Δ, waves considered marginally well-sampled and well-sampled, respectively (Doswell and Caracena 1988). For L=10Δ (Fig. 4), the rmse and correlation coefficient values indicate that the checkerboard pattern is represented best by the triangle method. The finite differencing method tends to produce poor divergence estimates in data void regions (Fig. 4b), a result consistent with Barnes (1994) and SD01. Although the rmse for the finite differencing method is less than that of the pentagon method, the pattern of the pentagon analysis is superior, as reflected by the correlation coefficients. For L=20Δ (Fig. 5), both the triangle and pentagon analyses appear to be reasonable representations of the analytic field, although the rmse and correlation coefficient values indicate that the triangle method is superior. Again, the finite differencing analysis is clearly the poorest of the three.

Several conclusions may be drawn from Fig. 6, which compares the performances of the three methods for various values of $c_{2,3}$. First, as the wavelength increases, the magnitudes of the errors for each of the methods decrease. Errors generally decrease by an order of magnitude or greater as the wavelength increases from L=5Δ to L=20Δ. Clearly, for a given observing network, longer wavelengths are sampled better than shorter wavelengths, thus reducing the analysis error.

Second, the rmse values have a "U"-shaped pattern. Generally, the lowest rmse values are associated with smoothing parameters falling in the range of $0.8 \leq c_{2,3} \leq 1.0$. However, at L=5Δ, slightly lower values of $c_{2,3}$ are required to minimize the rmses. When "excessively small" smoothing parameters are used (e.g. $c_{2,3} < 0.8$), an overfitting

of the observations yields relatively high rmse values. Alternately, with "excessively large" smoothing parameters (e.g. $c_{2,3} > 1.4$), the analysis is too smooth. In addition, increasingly large smoothing parameters cause boundary errors to creep towards the center of the domain; we have attempted to reduce these errors by restricting our verification domain as mentioned in the previous section.

Third, the average rmses of the pre-analyzed divergence estimates[1] at triangle centroids and interior stations of the pentagons (labeled "T" and "P", respectively, in Fig. 6) suggest that for $L=5\Delta$, the divergence estimates from the triangle method are superior to those of the pentagon method, whereas for larger wavelengths, the reverse is true. While all of the wavelengths considered are nonlinear, the $L=5\Delta$ wave is the most nonlinear with respect to station separation (Fig. 7). For this reason, intuition suggests that at this wavelength the pentagon method should be superior to the triangle method; clearly this is not the case. Apparently, for $L=5\Delta$, the linearity assumption over the relatively small triangles (average area = 6800 km^2) is superior to the quadratic assumption over the relatively large pentagons (average area = 31 000 km^2). For the larger wavelengths ($L \geq 10$), however, the reverse is true; namely, the linearity assumption over the relatively small triangles is *inferior* to the quadratic assumption over the relatively large pentagons. At the larger wavelengths, the nonlinearity with respect to station separation is much smaller (Fig. 7), but still large enough to allow superior analyses from the pentagon method.

When the analyses are performed using BOA3, the triangle method rmse minima are lower than the minima from both the finite differencing and pentagon methods for all

[1] These rmses refer to the average rmse values at the triangle centroids and interior stations of pentagons *before* the divergence estimates are analyzed to the grid via BOA3. Since errorless observations are used, these errors represent truncation errors.

wavelengths (Fig. 6). For the smaller wavelengths (L=5Δ and L=10Δ), the finite differencing method generally outperforms the pentagon method; for the larger wavelengths (L=15Δ and L=20Δ), the opposite is true. Interestingly, for all wavelengths, the rmses of the pentagon analyses are greater than the rmses of the pre-analyzed divergence estimates, whereas for wavelengths exceeding L=5Δ, the rmses of the triangle analyses are smaller than the rmses of the pre-analyzed divergence estimates. We believe this to be a consequence of two factors. First, the pentagon analysis involves only 155 divergence estimates, whereas the triangle analysis involves 361 divergence estimates; the more data involved in the analysis, the less the error. Second, because the signs of the errors of the triangle centroid divergence estimates are somewhat randomly distributed in space (Fig. 8), we believe that the analysis itself acts to reduce the effects of these errors. This is not true for the pentagon method, where the distribution of the signs of the errors generally corresponds to the pattern of the divergence field (Fig. 9).

Figure 10 indicates that for all wavelengths considered, larger triangle and pentagon areas are associated with larger truncation errors. Even a highly nonlinear field varies approximately linearly across small triangles such that the linearity assumption of the triangle method produces a reasonable estimate of the divergence. For larger triangles, the degree of nonlinearity of the flow within the triangle is larger as well, resulting in greater errors associated with the linearity assumption. A similar argument can be made for pentagon area versus error.

For the smallest wavelength considered (L=5Δ), even the smallest pentagons, with their less restrictive quadratic assumption, have truncation errors exceeding those of small triangles (Fig. 10a). Specifically, approximately 55 of the 361 triangles (~15%)

have truncation errors less than those of the smallest pentagons. However, as the wavelength increases, an increasingly large number of pentagons have truncation errors less than those of the smallest triangles. For example, Fig. 10b indicates that for any pentagon whose area is less than about 18 000 km^2, the quadratic assumption is better than the linearity assumption for all triangles. At L=20Δ, the quadratic assumption is better for any pentagon whose area is less than about 38 000 km^2 (Fig. 10d).

5. Summary and conclusions

This study has shown that for computing gridded fields of divergence from irregularly spaced wind observations, the triangle method consistently outperforms the finite differencing and pentagon methods when shape parameter values within the range generally recommended are used (e.g., c_1=1.75, $c_{2,3}$=1.0). For well-sampled waves, the pentagon analyses clearly are superior to those from the finite differencing method. An important difference between the three methods is that the triangle and pentagon methods apply an analysis scheme to estimates of divergence that are computed *directly* from the observations, whereas the finite differencing method computes divergence from gridded fields of the horizontal wind components. Doswell and Caracena (1988) show that this difference produces different analyses when the observations are irregularly distributed.

The pre-analyzed divergence estimates from the pentagon method are superior to those from the triangle method for wavelengths L≥10Δ, but are inferior for L=5Δ. We find this to be an interesting—if not counterintuitive—result since the nonlinearity of the flow with respect to the station separation is greatest at L=5Δ and the stated purpose of the pentagon method is to capture a portion of that nonlinearity. Apparently, the

linearity assumption over the relatively small triangles is superior to the quadratic assumption over the relatively large pentagons for $L=5\Delta$; the opposite is true at larger wavelengths. We have found that when the BOA3 scheme is applied to the irregularly spaced divergence estimates at the interior stations of pentagons, the rmse values increase. On the other hand, when the BOA3 scheme is applied to the irregularly spaced divergence estimates at the triangle centroids, the rmse values generally decrease. We consider this behavior to be a consequence of both the number of irregularly spaced divergence estimates (361 for the triangle method and 155 for the pentagon method) and the smoothing effect of the BOA3 scheme on the errors of the divergence estimates from the triangle method.

Acknowledgements. This work is supported by the National Science Foundation under Grant 0097651. We thank Daphne Zaras for coordinating the Research Experiences for Undergraduates program, and her assistant, Lance Maxwell, for his help throughout the summer. We wish to acknowledge Dr. Charles Doswell and Mr. Dave Watson for providing us the Delauney triangulation code, without which this work would be impossible. We thank Dr. Phillip Smith for his suggestions regarding the selection of pentagons. Finally, we appreciate the suggestions of Dr. David Schultz concerning the various drafts of this paper.

References

Achtemeier, G. L., 1987: On the concept of varying influence radii for successive corrections objective analysis. *Mon. Wea. Rev.,* **115,** 1760-1771.

-----, 1989: Modification of a successive corrections objective analysis for improved derivative calculations. *Mon. Wea. Rev.,* **117,** 78-86.

Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.,* **3,** 396-409.

-----, 1973: Mesoscale objective analysis using weighted time-series observations. NOAA Tech. Memo. ERL NSSL-62, National Severe Storms Laboratory, Norman, 41 pp. [NTIS COM-73-10781.]

-----, 1994: Applications of the Barnes object analysis scheme. Part I: Effects of undersampling, wave position, and station randomness. *J. Atmos. Oceanic Technol.*, **11,** 1433-1448.

Chien, H., and P. J. Smith, 1973: On the estimation of kinematic parameters in the atmosphere from radiosonde wind data. *Mon. Wea. Rev.,* **101**, 252-261.

Doswell, C. A., III, and F. Caracena, 1988: Derivative estimation from marginally sampled vector point functions. *J. Atmos. Sci.,* **45**, 242-253.

Ripley, B. D., 1981: *Spatial Statistics*. Wiley-Interscience, 252 pp.

Schaefer, J. T., and C. A. Doswell III, 1979: On the interpolation of a vector field. *Mon. Wea. Rev.*, **107,** 458-476.

Spencer, P. L., and C. A. Doswell III, 2001: A quantitative comparison between traditional and line integral methods of derivative estimation. *Mon. Wea. Rev.,* **129,** 2538-2554.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

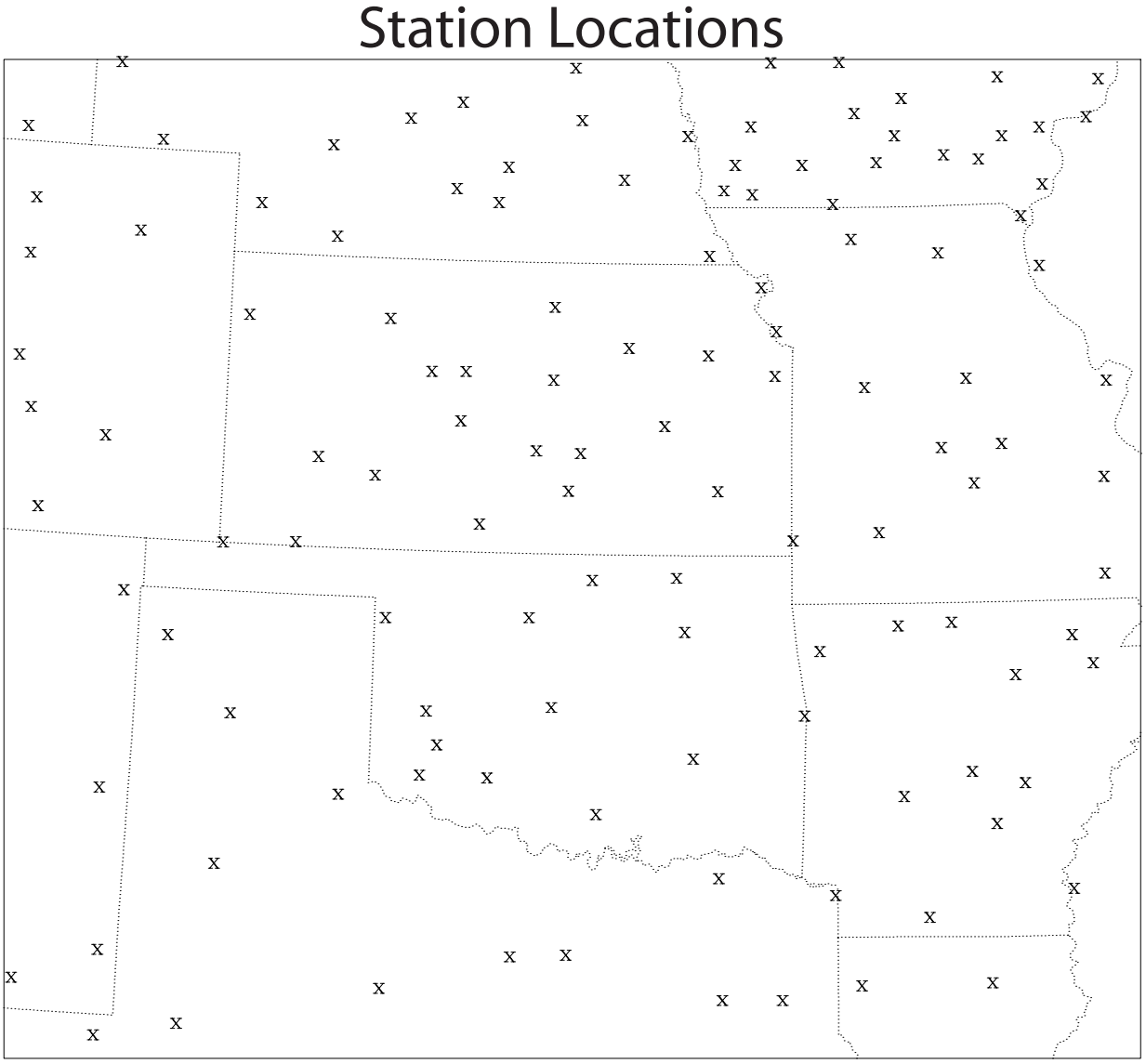

Fig. 1. Declustered surface station network considered in the analysis.

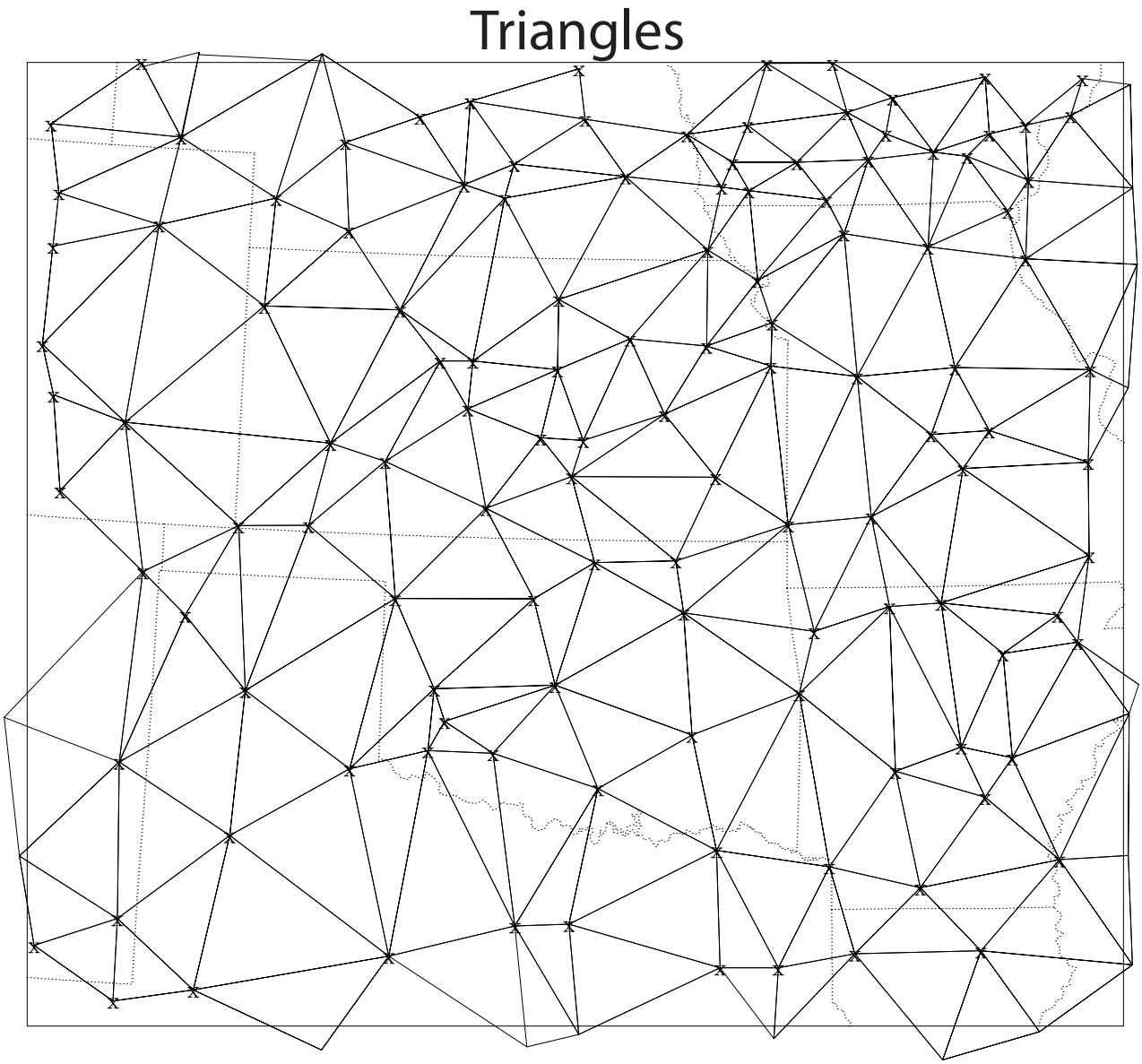

Fig. 2. The Delauney triangulation of the observation network shown in Fig. 1. For clarity, some of the triangles along the outer edges are not drawn.

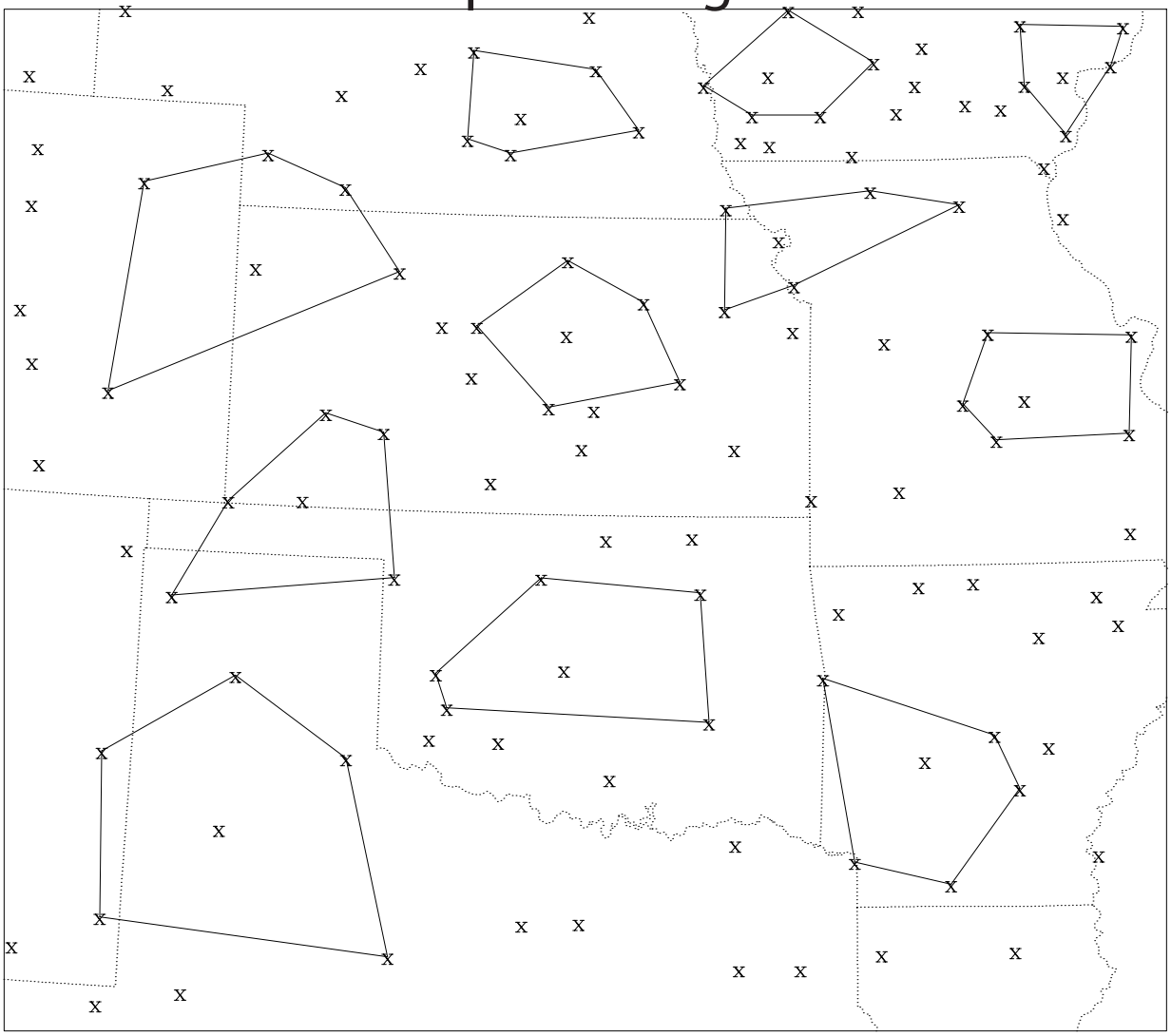

Fig. 3. A sampling of the pentagonal tessellation of the observation network.

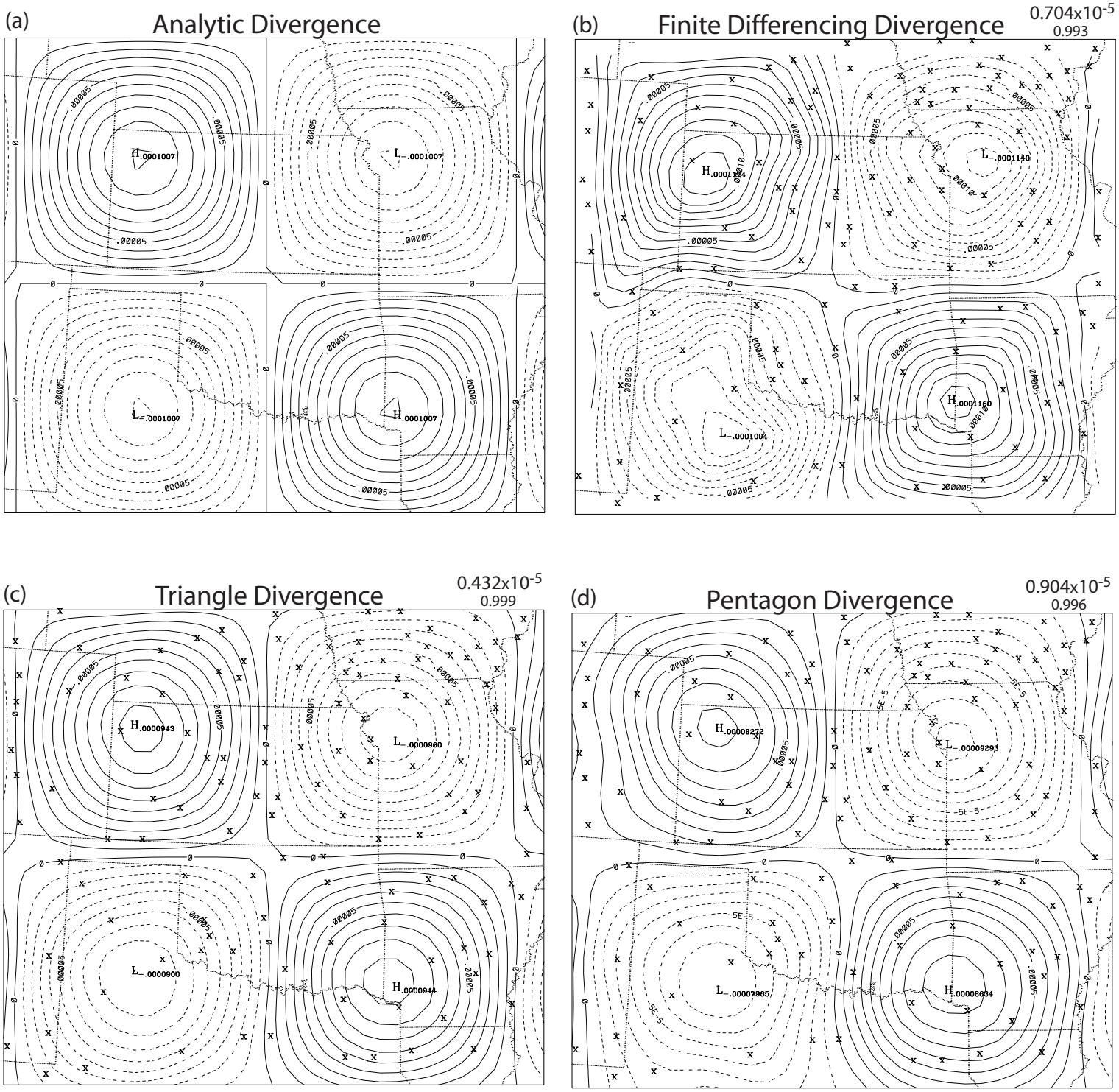

Fig. 4. Contour plots of divergence from the (a) analytic function, (b) finite differencing method, (c) triangle method, and (d) pentagon method for L=10Δ. The analysis parameters are c_1= 1.75 and $c_{2,3}$=1.0. For these plots, $\phi_x = \phi_y = 0$. In (b)-(d), observation locations are indicated by an "x". The rmses (s^{-1}) and correlation coefficients for the analyses are shown in the upper right corner of (b)-(d). The contour interval for all plots is the same.

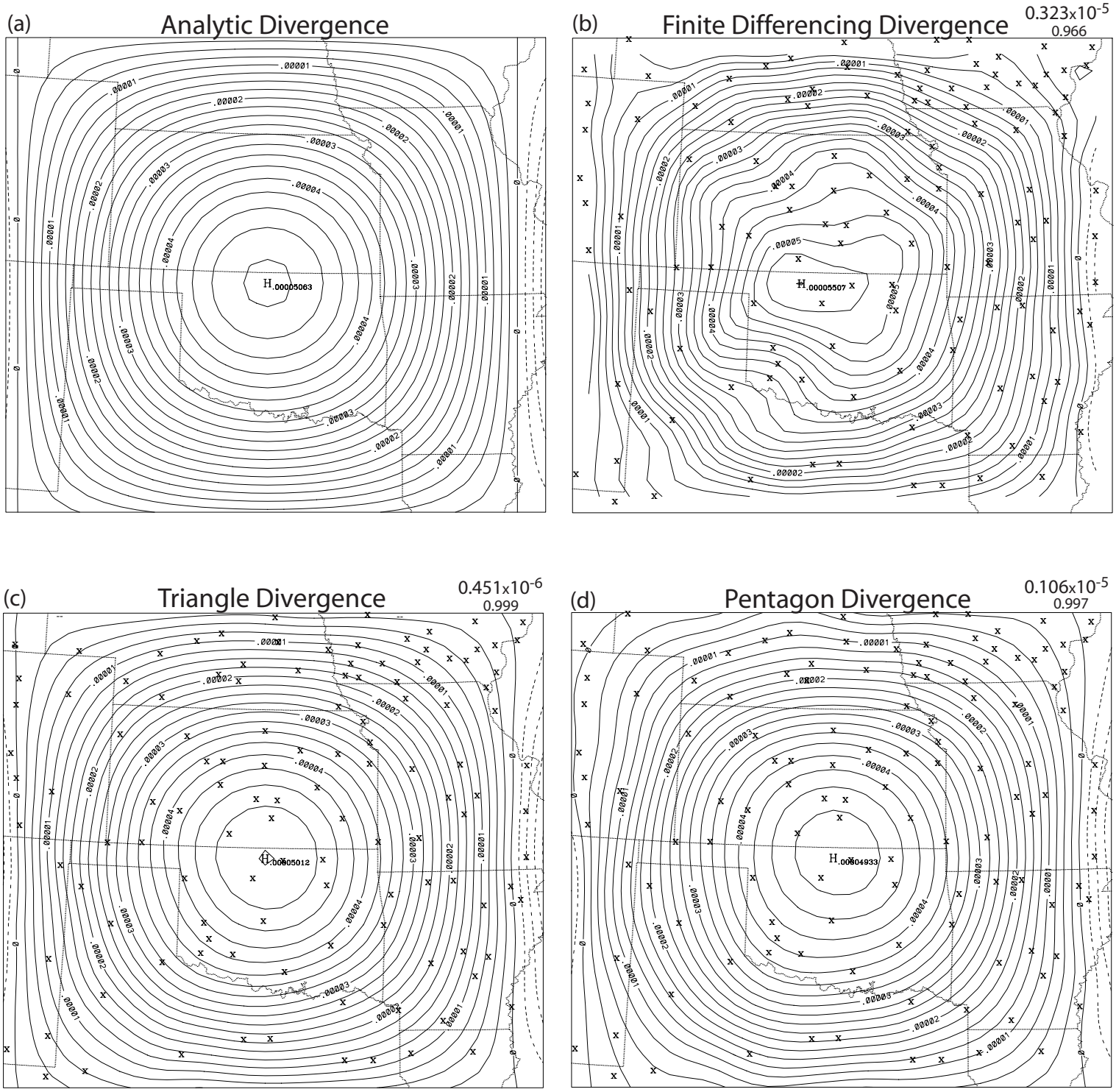

Fig. 5. Same as Fig. 4, except $L=20\Delta$, $\phi_x=-\pi/2$, $\phi_y=\pi/2$.

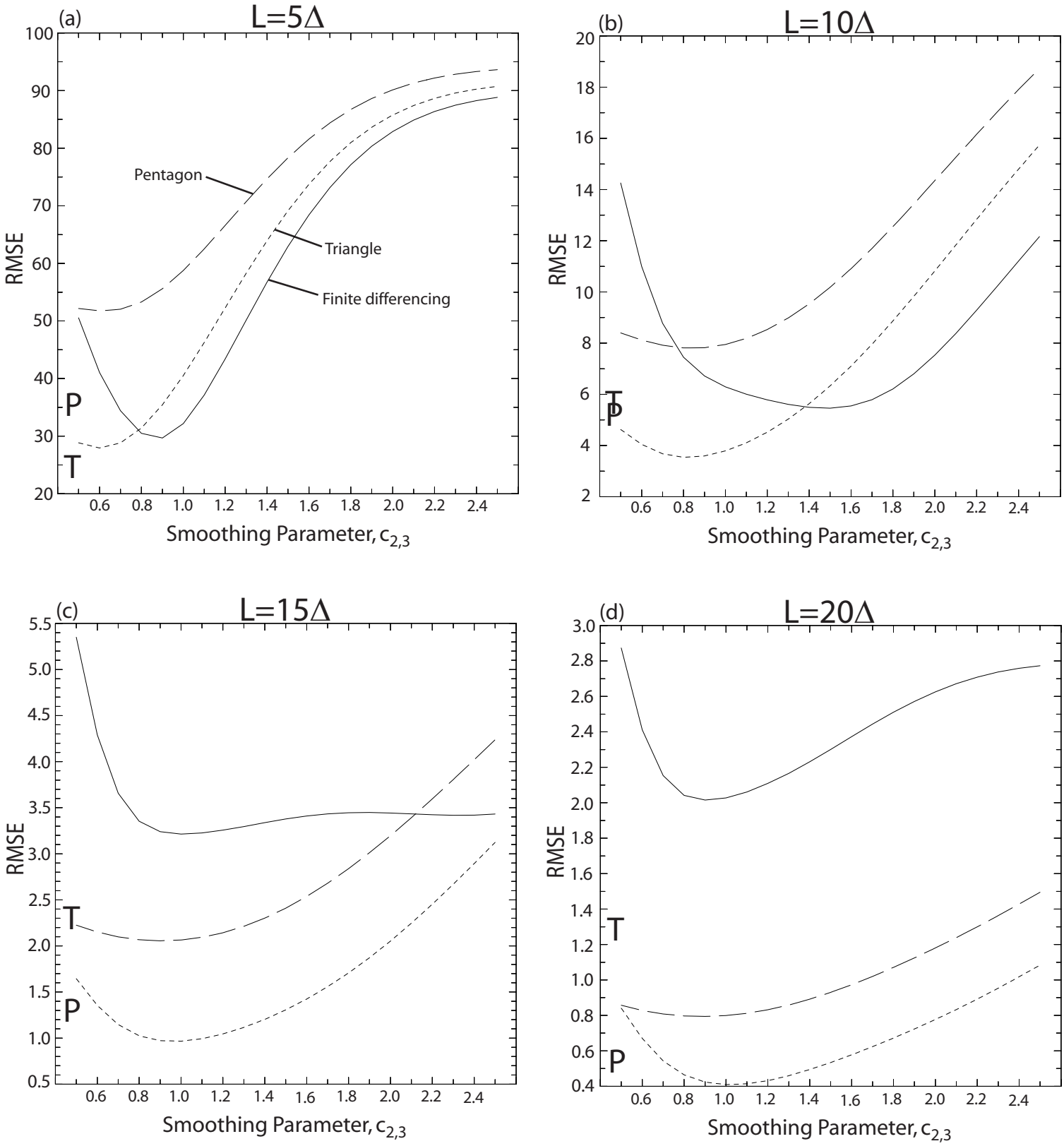

Fig. 6. Root mean square errors as a function of $c_{2,3}$ for the finite differencing (solid curves), triangle (short-dashed curves), and pentagon (long-dashed curves) methods for (a) L=5Δ, (b) L=10Δ, (c) L=15Δ, (d) L=20Δ. The "T" on each of the plots indicates the average rmse of the pre-analyzed divergence estimates at triangle centroids. The "P" represents the average rmse of the pre-analyzed divergence estimates at the interior stations of pentagons. For all curves, c_1=1.75. Each of these curves represents averages of nine analyses whose phase shifts (ϕ_x, ϕ_y) are: (-π/4,π/4), (0,π/4), (π/4,π/4), (–π/4,0), (0,0), (π/4,0), (–π/4,–π/4), (0,–π/4), (π/4,–π/4).

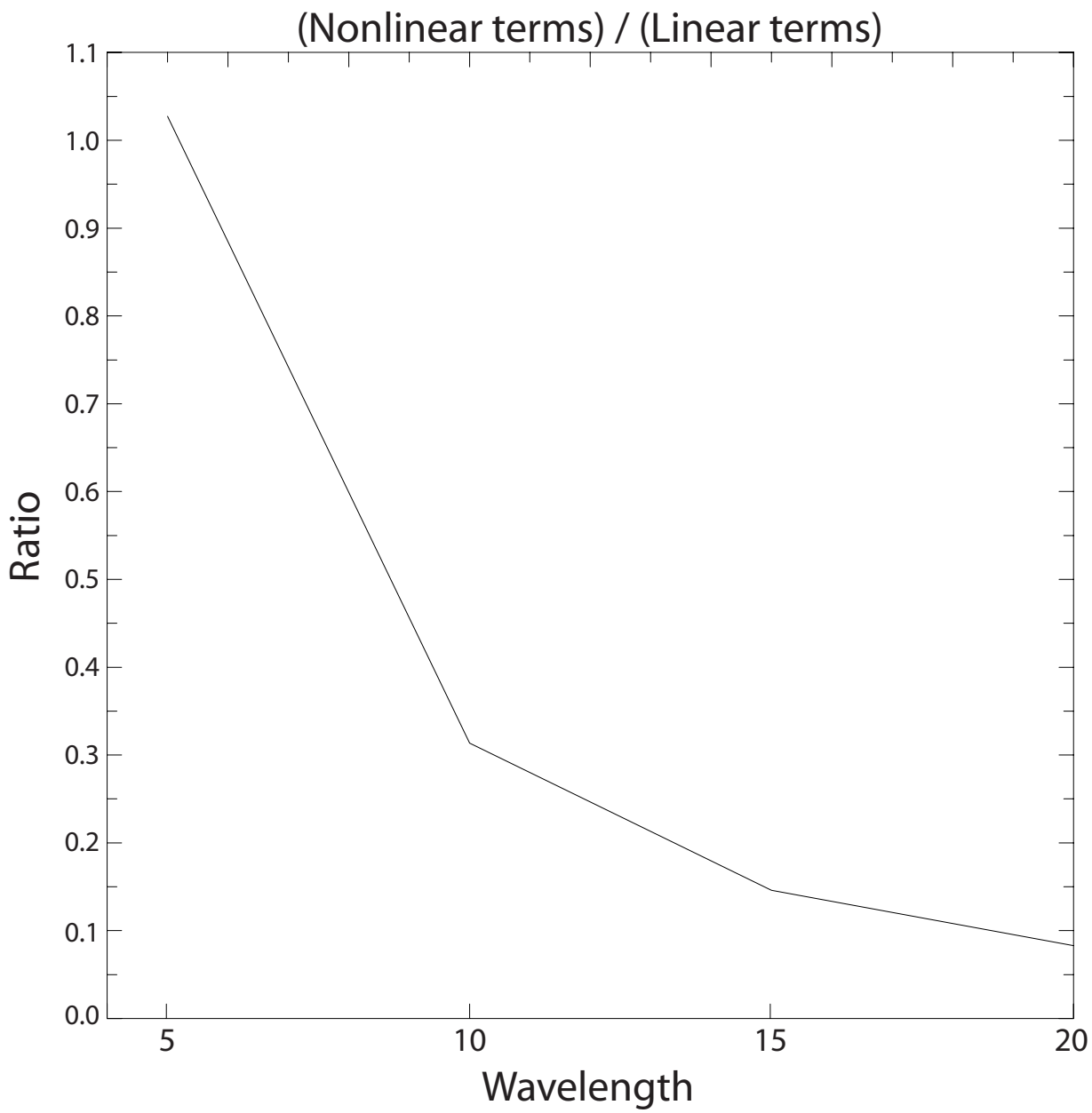

Fig. 7. Relationship between the wavelength and the ratio of the magnitude of the sum of the nonlinear terms to the magnitude of the sum of the linear terns from the quadratic Taylor series representation of the wind field. As in Fig. 6, the curve represents averages of nine sets of analyses.

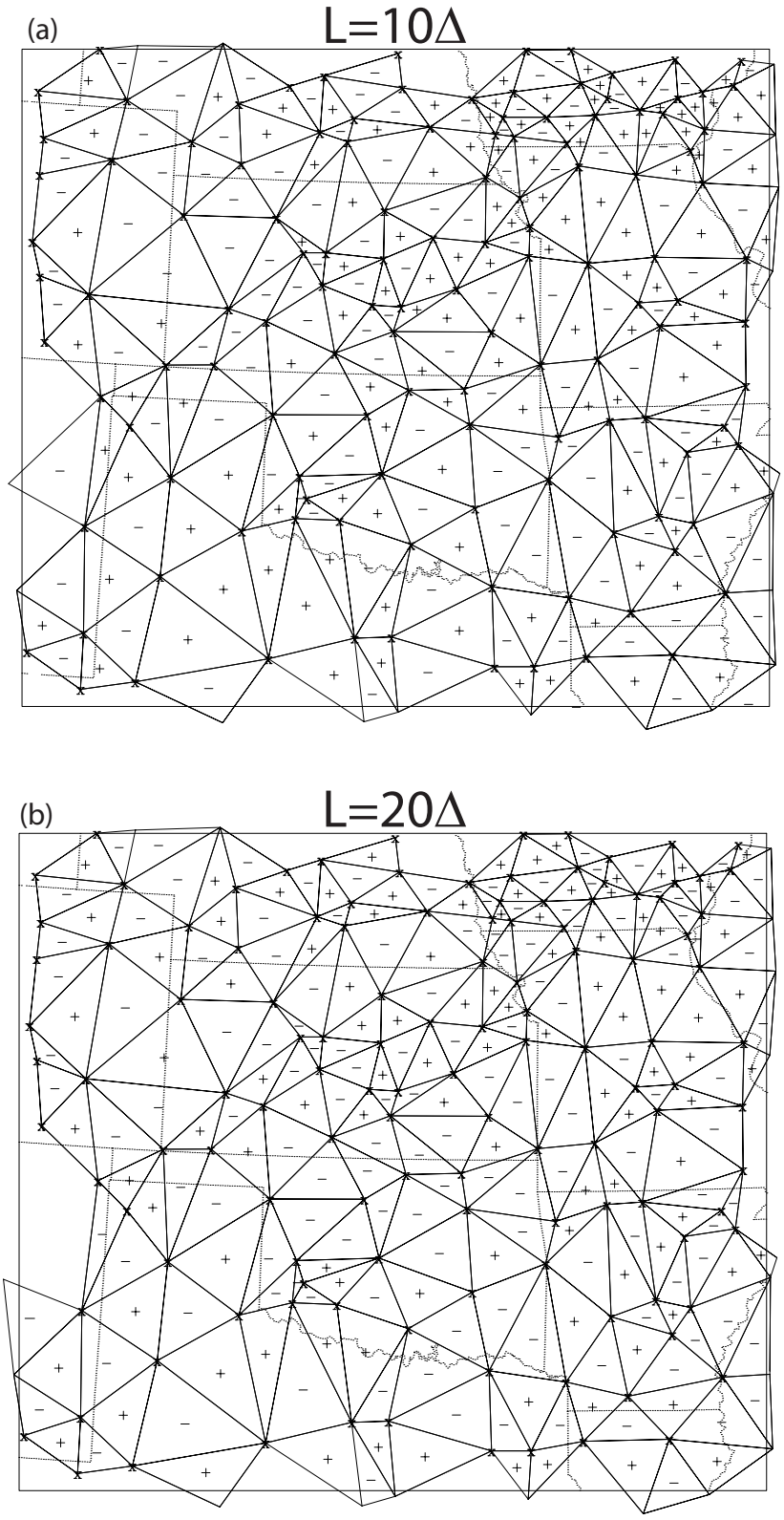

Fig. 8. Signs of the differences between the analytic divergence at triangle centroids and the divergence computed from the triangle method for (a) L=10Δ and (b) L=20Δ. Plus signs indicate that the computed divergence is greater than the analytic divergence. The analytic divergence used in (a) is the same as that presented in Fig. 4a and the analytic divergence used in (b) is the same as that presented in Fig. 5a.

Fig. 9. Signs of the differences between the analytic divergence at the interior stations of pentagons and the divergence computed from the pentagon method for (a) L=10Δ and (b) L=20Δ. Plus signs indicate that the computed divergence is greater than the analytic divergence. The analytic divergence used in (a) is the same as that presented in Fig. 4a and the analytic divergence used in (b) is the same as that presented in Fig. 5a.

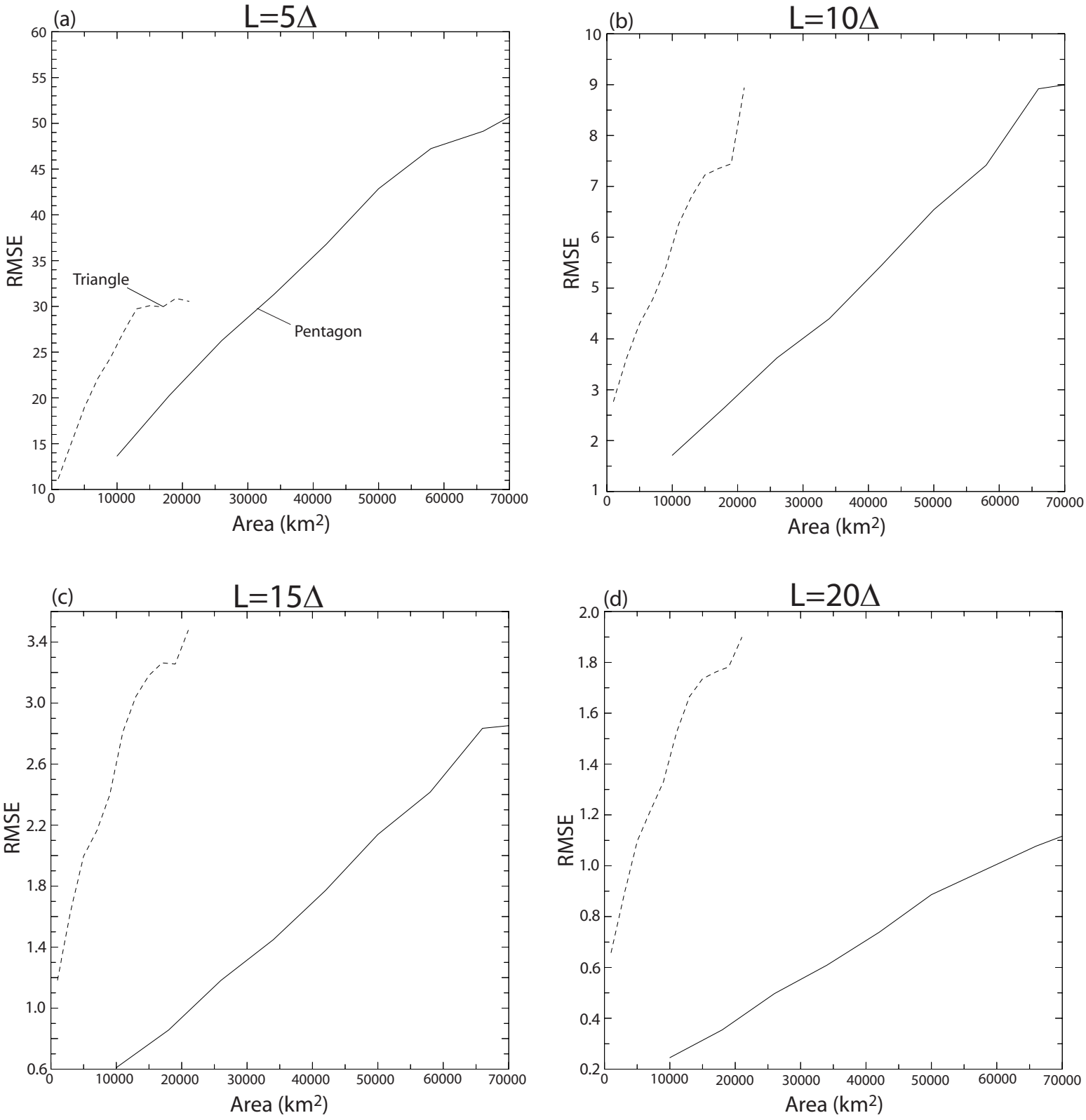

Fig. 10. Root mean square errors of the pre-analyzed divergence estimates from the triangle (dashed curves) and pentagon (solid curves) methods as a function of the area of the triangle or pentagon for (a) L=5Δ, (b) L=10Δ, (c) L=15Δ, (d) L=20Δ. The areas of the triangles are binned in 2000 km^2 increments. Errors of all triangles associated with each bin are averaged to create the dashed curves. A similar procedure is used for the pentagons, except the bin increment is 8000 km^2. As in Fig. 6, the curves represent averages of nine analyses.