

*Н. М. Оскорбин, И. Н. Ротанова, С. И. Суханов*

### **Создание ГИС-полигонов для оценки согласованности и точности пространственных данных в условиях Алтайского края\***

*N. M. Oskorbin, I. N. Rotanova, S. I. Sukhanov*

### **Setting up GIS Polygons to Assess Coordination and Accuracy of the Spatial Data in the Altai Krai**

Изучены проблемы оценки согласованности и точности пространственных данных, компьютерные базы которых создаются в настоящее время в ряде субъектов Российской Федерации, в том числе в Алтайском крае в рамках проектов разработки инфраструктуры пространственных данных. Для решения этой проблемы предлагается интервальный подход к анализу данных и совокупность ГИС-полигонов, которые выступают в качестве тестовых территорий.

**Ключевые слова:** базовые пространственные данные, методы анализа данных, прикладной интервальный анализ, геоинформационные системы (ГИС).

Современные методы анализа массивов данных находят применение в самых различных областях теоретических и прикладных исследований, в том числе в геоинформатике при обработке геодезических измерений, построении и коррекции карт и планов, в кадастровых системах различного назначения и при создании прикладных геоинформационных систем (ГИС). Процедуры анализа массивов пространственных данных включают их предобработку, проверку согласованности, полноты и степени пригодности для решения конкретных прикладных задач.

В статье рассмотрены проблемы оценки согласованности и точности существующих пространственных данных путем использования дополнительной геоинформации, которую предлагается получать на ГИС-полигонах, выступающих в качестве тестовых территорий. Методической основой решения этих задач являются результаты работ [1–6].

Следует отметить, что традиционно для решения задач анализа пространственных данных используются вероятностно-статистические методы: регрессионный, корреляционный, дисперсионный анализы, методы математической статистики, многомерного анализа данных и др. В последнее время интенсивно развивается нестатистический подход к анализу пространственных данных, основная идея которого состоит в замене вероятностной модели неопределенных факторов на интервальную модель [6].

The article focuses on estimating coordination and accuracy of the spatial data, which form a framework of computer data in some federal subjects of Russia including Altai Krai, within the project of developing a spatial data infrastructure (SDI). To solve the problem we use interval approach to analyze the data and set of GIS polygons which are selected to be testing territories.

**Key words:** basic spatial data, data analysis techniques, applied interval analysis, geographic information system (GIS).

С формальной точки зрения, при отказе от вероятностного описания неопределенности происходит потеря информации, однако за счет получения устойчивых оценок, возможности привлечения априорной информации, эффективного выявления совокупностей противоречивых данных и знаний эта потеря в ряде прикладных задач компенсируется. При интервальном оценивании усложняются и численные методы, которые в интервальном анализе применяются для решения сложных экстремальных задач, однако они могут базироваться на современных пакетах прикладных программ.

Рассмотрим методику оценивания точности двумерных и трехмерных картографических данных и геодезических измерений с использованием математических моделей пространственного положения географических объектов и методов интервального анализа. В работе [6] проведено исследование возможностей применения интервального анализа к задачам обработки геоданных и показано, что для его использования необходимы математические модели пространственного положения геообъектов как для получения точечных оценок, так и для их локальных перемещений. Задача оценки согласованности и точности картографического отображения двумерных и трехмерных геоданных формализуется следующим образом.

Пусть на исследуемой территории выделена совокупность геообъектов, индексы которых принад-

\* Работа подготовлена при финансовой поддержке РФФИ (проект 12-07-98012 — р\_сибирь\_a — Региональный конкурс СИБИРЬ).

лежат множеству  $J = \{1, 2, \dots, N\}$ . Пространственное положение объекта  $j \in J$  определяется в некоторой (расчетной) системе координат массивом точек  $M_j = \langle (X_i^j, Y_i^j, Z_i^j), i=1, \dots, n_j \rangle$ , истинные значения которых априори неизвестны. Здесь  $n_j$  — число точек, достаточное для полного описания формы и пространственного положения объекта  $j \in J$ .

Информация, которой располагает исследователь для построения оценки  $\hat{M}_j, j \in J$  истинного положения  $j$ -го объекта и для оценки всего массива геоданных  $\hat{M} = \langle \hat{M}_j, j \in J \rangle$ , включает:

- 1) знание формы и размеров каждого исследуемого геобъекта;
- 2) априорные качественные и количественные сведения о положении геобъектов и их пространственных отношениях;
- 3) результаты геодезических измерений  $\tilde{M}_j, j \in J$  координат объекта (этот массив по структуре, составу элементов и системе координат в общем случае не совпадает с массивом  $M_j, j \in J$ );
- 4) полное описание условий геодезических измерений и сбора картографических данных, включая оценки погрешности полевых измерений или «сколки» их значений с карты.

Исходная информация включает в себя как данные, так и знания и характеризуется значительным объемом. Обработка этой информационной совокупности может быть выполнена методами интервального анализа, которые позволяют получать не только точечные оценки  $\hat{M}_j, j \in J$  пространственного положения каждого объекта, но и описывать каждое множество  $Q_j$  (информационное множество) согласованных значений массива  $M_j (M_j \in Q_j), j \in J$ . В работах [2–4] рассмотрена структура множества  $Q_j$  с использованием понятия «порция информации». Пусть  $Q_j^v, v=1, \dots, L_j$  — элементарные информационные множества, построенные на основе групп знаний и данных, указанных в пунктах 1, 3, 4 выше приведенного списка. Тогда

$$Q_j = \bigcap_{v=1}^{L_j} Q_j^v. \quad (1)$$

Здесь  $L_j$  — количество выделенных групп данных, которое включает число измеренных координат массива  $\tilde{M}_j$  и число формализованных условий на основе априорной информации,  $j \in J$ .

Пусть также  $Q_\Sigma^v, v=1, \dots, L_\Sigma$  — совокупность «порций информации», построенных с использованием данных и знаний пункта 2. Множество  $Q_\Sigma^v$  введено для описания взаимосвязей между отдельными объектами и определяет совместное пространственное положение объектов, т. е.  $M \in Q_\Sigma^v$ .

В том случае, когда исходная информация не противоречива, имеют место включения:  $M_j \in Q_j, j \in J$ ,

$M \in Q_\Sigma = \bigcap_{v=1}^{L_\Sigma} Q_\Sigma^v$ , а все записанные множества являются

непустыми. Однако проверить эти условия не представляется возможным. На практике ограничиваются проверкой согласованности используемых данных и знаний.

*Определение 1.* Информационным множеством пространственного положения всех геобъектов  $j \in J$  называется множество  $Q$ , такое что:

$$Q = \bigcup_{j \in J} Q_j \cap Q_\Sigma. \quad (2)$$

*Определение 2.* Информация о пространственном положении геобъектов согласована, если  $Q$  — непустое множество ( $Q \neq \emptyset$ ).

Исследование причин несогласованности информационных множеств показывает, что имеющейся информации в созданной базе данных и знаний недостаточно для выявления противоречивых групп «порций информации». Пусть, например, множество  $Q = \emptyset$  и исключение порции  $Q_j^1$  обеспечивает условие согласованности оставшихся «порций информации», т. е. выполнение условия:  $Q_j \setminus Q_j^1 = \bigcap_{v=2}^{L_j} Q_j^v \neq \emptyset$ . Можем ли мы

утверждать, что «порция»  $Q_j^1$  не соответствует реальности, т. е.  $M_j \notin Q_j^1$ ? Оказывается, что в рамках нашей формализации и имеющейся информации этого утверждать нельзя. Необходима дополнительная географическая информация, которая может подтвердить, что  $M_j \notin Q_j^1$ , тогда  $Q_j^1$  исключается из базы информационной системы, или выявить, что  $M_j \in Q_j^1$ . В последнем случае анализ необходимо провести для всех оставшихся «порций информации». Из приведенного теоретического анализа следует необходимость проведения специальных геодезических измерений для проверки согласованности и точности пространственных данных. Очевидно, что полного обследования всех противоречивых пространственных данных провести невозможно. Однако можно исследовать закономерности искажений имеющихся пространственных данных на тестовых территориях. Такие территории могут быть названы ГИС-полигонами [2].

Продолжим описание методики анализа пространственных данных. На практике при обработке данных необходимо выявить минимальный набор «порций», для которых нужен дополнительный анализ. Для решения этой задачи методами прикладного интервального анализа можно использовать следующий алгоритм. Пусть  $k_j^v$  — параметры, значения которых «расширяют» или «сужают» «порции»  $Q_j^v$ . Определим  $Q_j^v(k_j^v)$  так, чтобы выполнялись условия:

$$\begin{aligned} Q_j^v(1) &= Q_j^v; & ( ) & ( ), & , & , & , & . \\ Q_j^v(k_{j_1}^v) &\subset Q_j^v(k_{j_2}^v), & k_{j_1}^v &< k_{j_2}^v, & v=1, \dots, L_j, & j \in J. \end{aligned} \quad (3)$$

Аналогично введем параметры  $k_{\Sigma}^v, v=1, \dots, L_{\Sigma}$ . Пусть  $k_j = (k_j^v, v=1, \dots, L_j), k_{\Sigma} = (k_{\Sigma}^v, v=1, \dots, L_{\Sigma})$ . Тогда  $k = (k_j, k_{\Sigma}, j \in J)$ .

Определим расширяемое или сужаемое информационное множество, модифицируя (2):

$$Q(k) = \bigcup_{j \in J} Q_j(k_j) \cap Q_{\Sigma}(k_{\Sigma}), \quad (4)$$

и отыщем  $k^*$  решением следующей экстремальной задачи:

$$\min_{M, k \geq 1} \left\{ \sum_{j \in J} \sum_{v=1}^{L_j} k_j^v + \sum_{v=1}^{L_{\Sigma}} k_{\Sigma}^v \mid M \in Q(k) \neq \emptyset \right\}. \quad (5)$$

Равенство единице всех коэффициентов расширения  $k^*$ , полученных при решении задачи (5), означает согласованность данных. В противном случае можно выявить группу «порций информации», которые препятствуют согласованности, и дополнительно проверить их достоверность. Для скорректированной информационной совокупности следует также решать задачу (5).

Для исследования альтернативного способа согласования данных можно предложить модификацию задачи (5), в которой все коэффициенты расширения одинаковы и равны  $k$ , тогда  $k^*$  следует отыскивать из условия:

$$k^* = \min_{M, k \geq 1} \{k \mid M \in Q(k) \neq \emptyset\}. \quad (6)$$

Результаты решения задачи (6) в случае несогласованности данных позволят провести ранжирование «порций», препятствующих согласованию.

Пусть исходная информация согласована, т. е.  $Q \neq \emptyset$ . Рассмотрим задачи оценки точности пространственных координат. Зададим интересующую нас скалярную характеристику  $h$  совокупности геообъектов как функцию истинных координат  $h = H(M)$ . Таких характеристик может быть много. Оценки  $\hat{h}$  и  $\hat{M}$  можно найти при общем коэффициенте расширения решением следующей задачи (оценки центра неопределенности):

$$\hat{h} = H(\hat{M}); \quad \hat{M} = \arg \min_{M, k \geq 0} \{k \mid M \in Q(k)\}. \quad (7)$$

Другим способом нахождения оценок характеристик  $h$  совокупности геообъектов и погрешности этой оценки  $\Delta h$  является решение двух экстремальных задач (в условиях существования экстремумов):

$$\underline{h} = \min_{M \in Q} H(M); \quad \bar{h} = \max_{M \in Q} H(M). \quad (8)$$

Тогда

$$\hat{h} = \frac{\underline{h} + \bar{h}}{2}; \quad \Delta h = \frac{\bar{h} - \underline{h}}{2}. \quad (9)$$

Анализ задач (5) — (8) при исследовании совокупностей пространственных данных показывает, что, в общем случае, они относятся к многоэкстремальным задачам большой размерности. Для их существенного упрощения предложено использовать математическую модель пространственного положения

каждого геообъекта  $j \in J$ , координаты которого получаются при формировании информационного множества с использованием преобразования следующего вида:

$$\begin{pmatrix} X_i^j \\ Y_i^j \\ Z_i^j \end{pmatrix} = m_j \begin{pmatrix} \cos(\alpha_j) & \sin(\alpha_j) & 0 \\ -\sin(\alpha_j) & \cos(\alpha_j) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_i^l \\ Y_i^l \\ Z_i^l \end{pmatrix} + \begin{pmatrix} X_0^j \\ Y_0^j \\ Z_0^j \end{pmatrix}, \quad (10)$$

где  $m_j, \alpha_j, X_0^j, Y_0^j, Z_0^j$  — параметры преобразования, выбранные при решении задач (5) — (8);  $(X_i^j, Y_i^j, Z_i^j)$  — координаты геообъекта  $j \in J$  в расчетной системе координат;  $(X_i^l, Y_i^l, Z_i^l)$  — координаты характерных точек выбранного геопримитива  $I$ ;  $i (i=1, \dots, L_j)$  — индекс характерной точки выбранного геопримитива.

Отметим, что изменением параметров преобразования в формуле (10) реализуются локальные перемещения геообъектов, необходимые для получения интервальных оценок исследуемых характеристик.

Рассмотрим задачу формирования системы ГИС-полигонов в Алтайском крае, которая способствовала бы оценке согласованности и точности БПД. Мы предлагаем создавать такую систему по всем семи категориям земельных ресурсов, в том числе по землям населенных пунктов, сельскохозяйственного назначения, водного, лесного фондов, по землям особо охраняемых территорий. Такие полигоны могут быть размещены в районах края, в которых востребованы пространственные данные. В качестве методического подхода к проектированию таких полигонов можно использовать результаты работы [2], в которой описан опыт формирования базы пространственных данных ГИС-полигона, созданного на территории Барнаула.

Для проведения исследования согласованности и точности пространственных данных топографических карт, космических снимков и др. был выбран участок, расположенный в северо-западной части Барнаула, который включает территорию тепличного хозяйства, магистраль автомобильной дороги, железнодорожные пути, участок теплосети. Выбранная территория насыщена опознаваемыми точечными объектами (пересечение линейных объектов, характерные точки обочины магистральной автомобильной дороги, угловые точки границ земельных участков и зданий), линейными объектами и полигонами. На ней расположены пахотные земли (сельхозугодия), естественные ландшафты, примыкающие к защитной лесополосе, земельные участки, поставленные на кадастровый учет. Размеры выбранной территории (840x700 м) позволяют с достаточной точностью и полнотой исследовать пространственные данные, используемые в Барнауле. В качестве картографического материала при формировании картографических данных ГИС-полигона были выбраны разномасштабные карты (масштабы М 1: 10000, М 1: 2000,

М 1: 500), которые выступали одновременно объектами исследований. Участки растровых карт были предварительно обработаны с использованием программы ArcView GIS 3.2a, пересчетом их в местную систему координат Барнаула.

Все полевые работы проводились в сентябре 2009 г. Координаты объектов, легко идентифицируемых на картах и на местности, были определены с высокой точностью благодаря совместному применению спутниковой системы GPS Trimble 5700 и тахеометрических измерений. Измерения проводились двумя приемниками Trimble по схеме: база — подвижный приемник, совместно с тахеометрическими измерениями. Вычисление и уравнивание координат точек выполнено в программе «Trimble Geomatics Office» в местной системе координат Барнаула. Модуль погрешности взаимного места определения точек съёмочного обоснования не превышает  $\pm 0,05$  м.

В результате использования комплексной картографической и геодезической информации сформирована база геоданных, которая послойно представлена в ArcView GIS 3.2a в виде самостоятельной геоинформационной системы. Эта база использована для апробации математических моделей, алгоритмов и программ интервальной оценки точности пространственных данных.

В работе [2] в качестве базовых задач оценивания рассматриваются:

1. Оценка точности растровой карты с использованием метода центра неопределенности.
2. Интервальная оценка смещения высот топографической карты по данным GPS измерений.
3. Сглаживание массива данных рельефа с использованием сплайновых поверхностей.

При решении этих задач использовалось прямое и обратное преобразование геодезических данных, полученных в разных системах координат [5].

Первая из базовых задач анализа пространственных данных непосредственно направлена на оценку точности растровых и векторных карт и космических снимков. Опыт ее решения представлен в работе [2]. Рассмотрим методику и результаты исследования на примере оценки смещения высот растровой карты с использованием GPS измерений. Исследования проводились по следующей схеме. Предполагается, что рельеф растровой карты в каждой точке  $z_i^R$  смещен относительно истинного значения  $z_i^I$  на линейно зависящую от координат величину, которая представляется в интервальной форме и включает в себя

погрешность «сколки» и погрешность изготовления карты

$$z_i^R - z_i^I = a_0 + a_1 x_i + a_2 y_i, \quad i \in 1, \dots, N. \quad (11)$$

Общее число измерений составляло  $N = 110$ . Поиск коэффициентов проводился методом центра неопределенности решением задачи (7). В качестве независимых измерений  $z_i^I$  использовались данные высотных отметок, полученные с помощью системы спутниковой навигации GPS Trimble 5700. Погрешности GPS измерений  $\varepsilon^{GPS}$  не превышают 0,05 м, а погрешность «сколки»  $\varepsilon^R$  оценена равной 0,24 м. «Сколка» высот по растровой карте проводилась в точках установки GPS-приемника по ближайшим изолиниям исследуемой карты.

Введем информационное множество допустимых значений параметров в (11):

$$Q(k) = \left\{ (a_0, a_1, a_2) \left| \begin{aligned} z_i^{GPS} - z_i^R - k^R \varepsilon^R - k^{GPS} \varepsilon^{GPS} \leq \dots \\ \dots \leq a_1 + a_1 x_i + a_2 y_i \leq z_i^{GPS} - z_i^R + k^R \varepsilon^R + k^{GPS} \varepsilon^{GPS} \end{aligned} \right. \right\},$$

$$k = (k^{GPS}, k^R), \quad k^{GPS} = 1, \quad k^R = 1, \quad i \in 1, \dots, N.$$

При анализе согласованности информации гипотеза о нулевых значениях всех коэффициентов зависимости (11) не подтвердилась, что можно объяснить либо техногенными нарушениями рельефа, либо методическими погрешностями построения исследуемой карты на территории полигона.

Смещения рельефа в угловых точках территории полигона получены путем решения многоэкстремальной задачи на множестве  $Q(k)$  с целевой функцией (11). Соответствующие данные представлены в таблице.

Смещение рельефа в угловых точках ГИС-полигона, м

Угловые точки	$x_i$	$y_i$	$z_i^R - z_i^I$
Северо-запад	14017,07	-2642,70	0,76
Северо-восток	14017,07	-1645,57	0,42
Юго-восток	13633,36	-1645,57	0,28
Юго-запад	13633,36	-2642,70	0,62

Результаты расчетов показывают, что систематическое смещение истинных отметок высот по отношению к рельефу на карте масштаба 1:500 на территории полигона составляет в среднем 0,52 м.

На основании проведенных исследований можно рекомендовать методику прикладного интервального анализа и систему ГИС-полигонов для оценки согласованности, точности и корректирования пространственных данных при создании ИПД Алтайского края.

## Библиографический список

1. Кошкарев А. В. Эффективное управление пространственными метаданными и геосервисами в инфраструкту-

рах пространственных данных // Пространственные данные. — 2008. — № 1.

2. Оскорбин Н. М., Суханов С. И. Создание полигона для оценки точности имеющихся растровых карт и космических снимков высокого разрешения // Известия АлтГУ. — 2011. — № 1.

3. Оскорбин Н. М., Суханов С. И., Федин Л. Ю. Сглаживание массива данных рельефа с использованием сплайновых поверхностей // Известия АлтГУ. — 2010. — № 1.

4. Суханов С. И. Оценка точности растровой карты с использованием метода центра неопределенности // Известия АлтГУ. — 2010. — № 1.

5. Оскорбин Н. М., Суханов С. И. Оценка параметров формул прямого и обратного преобразования пространственных координат // Геодезия и картография. — М., 2011. — № 6.

6. Оскорбин Н. М., Жилин С. И., Суханов С. И. Интервальный подход к оценке согласованности и точности геоданных // Геодезия и картография. — М., 2011. — № 11.