

УДК 551.3

ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ ДАННЫХ МОНИТОРИНГА ЭКЗОГЕННЫХ ПРОЦЕССОВ, ПОЛУЧЕННЫХ ИЗ РАЗНОРОДНЫХ ИСТОЧНИКОВ

Шарапов Р.В.

*Муромский институт (филиал) ФГБОУ ВПО «Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»,
Муром, e-mail: info@vanta.ru*

В работе рассматриваются вопросы использования данных, полученных из разнородных источников. Такие данные могут получаться по различным методикам, с использованием разных приборов, могут иметь разную точность и погрешность измерений, а также содержать ошибки. Все это существенно усложняет их совместное использование и может приводить к значительным проблемам. Предлагается вместо применения непосредственных значений показателей, взятых из того или иного источника, использовать нормированные (преобразованные к одинаковым единицам измерения) тройки «значение, погрешность, шаг измерения». Это позволяет обеспечивать достаточную адекватность данных, собранных из различных источников, фактическому состоянию и осуществлять более точную оценку динамики процессов. Для сравнения данных предлагается учитывать «зону согласия», вхождение в которую обеспечивает непротиворечивость значений, полученных от различных источников.

Ключевые слова: мониторинг, экзогенные процессы, согласование, данные, источник информации, ошибка, погрешность, точность

FEATURES OF USING MONITORING DATA OF EXOGENOUS PROCESSES, WHICH ARE DERIVED FROM DIFFERENT SOURCES

Sharapov R.V.

Murom Institute of Vladimir State University, Murom, e-mail: info@vanta.ru

The paper deals with the use of data obtained from different sources. Such data may be obtained by various methods, using various devices, can have different accuracy and precision, can contain errors. All this greatly complicates their combined use and can lead to significant problems. It is proposed instead of using direct indicator values taken from one source or another, to use normalized (converted to the same units of measurement) triples «value, accuracy, measuring step». This solution enables us to provide sufficient adequacy of data, collected from various sources, to the practical situation and implement a more accurate assessment of the dynamics of processes. For comparison, data are encouraged to consider the «zone of agreement», the entry into which ensures not contradictory values obtained from various sources.

Keywords: monitoring, exogenous processes, conformity, data, data source, error, accuracy, precision

В современном научном мире нередко можно встретить случаи, когда при проведении наблюдений над одними и теми же объектами различные исследователи получают данные, существенно отличающиеся друг от друга [1]. Не является исключением и мониторинг экзогенных процессов (например, карстовых и суффозных) [4, 7, 8]. Это связано с рядом причин. Во-первых, разные исследователи при наблюдении за одними и теми же параметрами могут использовать различные методики. Во-вторых, могут использоваться различные приборы, отличающиеся точностью и погрешностью измерений. В-третьих, при проведении исследований могут возникать ошибки вследствие воздействия помех, неправильной работы оборудования и человеческого фактора [6].

Возникает проблема согласования и обработки данных, полученных из различных источников (исследователей). Особо остро проблема стоит в том случае, когда разные источники дают неполные данные (с разной

периодичностью, с наличием пропусков за определенные периоды и т.д.).

Цель работы – рассмотреть особенности согласования и обработки данных, полученных из различных источников при проведении мониторинга экзогенных процессов.

Несо согласованность данных

В 2013 году в ходе выполнения гранта РФФИ № 13-07-97510 проводились исследования влияния подземных вод на развитие экзогенных процессов [9]. Была проведена оценка состояния поверхностных и подземных вод округа Муром. В качестве основы для исследований были использованы данные, непосредственно полученные на кафедре техносферной безопасности МИ ВлГУ, данные Санитарно-гигиенического мониторинга Муромского района, сведения Роспотребнадзора и МУП «Водопровод и канализация». При анализе данных возникли некоторые сложности. Оказалось, что данные из разных источников плохо стыкуются друг с другом.

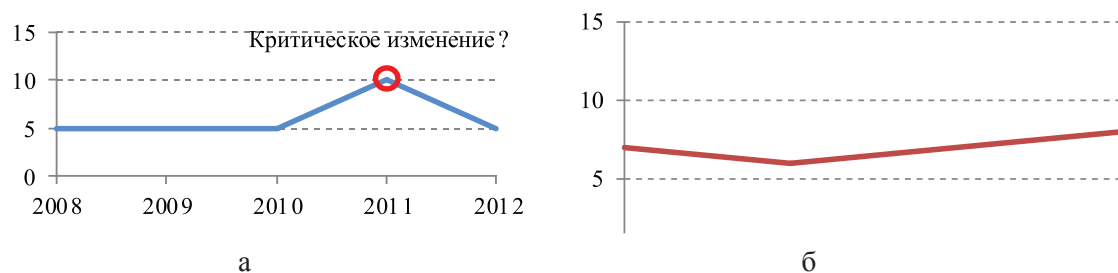


Рис. 1. Значения одного из параметров наблюдений за экзогенными процессами по данным различных источников:
а – источник 1; б – источник 2

На рис. 1, а приведено изменение показателя «Цветность» воды (измеряемого в градусах) по данным одного из источников. Как можно заметить, в 2011 году наблюдается существенный рост показателя (в 2 раза). Это может свидетельствовать о неких критических изменениях в режиме подземных вод и их влиянии на развитие экзогенных процессов. Однако при детальном анализе оказывается, что измерение параметра осуществляется с шагом 5 градусов. Таким образом, фактически произошло изменение на один шаг измерения. Анализ данных другого источника (рис. 1, б) для этого же показателя продемонстрировал иную картину. Шаг измерения этого источника составляет 1 градус. Как можно заметить, резких изменений показателя в 2011 году не наблюдается.

Фактически колебания составили 1 градус (с 7 в 2010 году до 8 в 2011). Таким образом, округления значений в первом случае при большом шаге измерений могут привести к ошибочным выводам.

Согласование данных

Пусть данные получают из двух источников информации D_1 и D_2 . $D_1 = \{d_1^1, d_2^1, d_3^1, \dots, d_n^1\}$, $D_2 = \{d_1^2, d_2^2, d_3^2, \dots, d_n^2\}$, где d_i^1 – данные наблюдений источника D_1 за i -й период, d_i^2 – данные наблюдений источника D_2 за i -й период (см. рис. 2). На практике данные источников за i -й период могут не совпадать друг с другом, т.е. $d_i^1 \neq d_i^2$ [3].

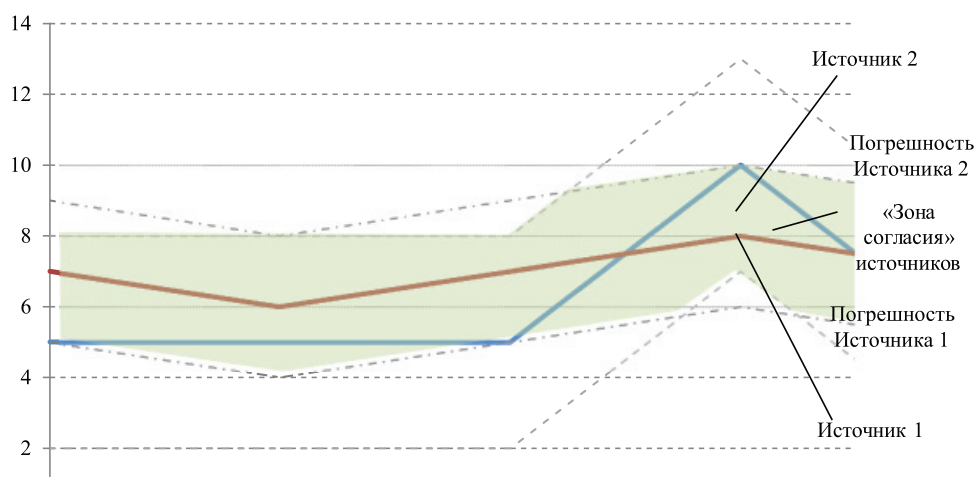


Рис. 2. Данные, полученные из разных источников

Каждый источник информации может иметь свои погрешности измерений (вызванные используемым оборудованием и методами исследования) [2]. Обозначим погрешность измерений источника D_1 как Δ_1 , погрешность источника D_2 как Δ_2 . Тогда фактическое значение параметров от источника D_1 будет лежать в диапазоне

$[d_i^1 - \Delta_1; d_i^1 + \Delta_1]$ («зоне погрешности»), от источника D_2 в диапазоне $[d_i^2 - \Delta_2; d_i^2 + \Delta_2]$. Другими словами, для каждого наблюдения можно задать как полученное значение, так и диапазон, в пределах которого могут быть отклонения этого значения от реальной (фактической) величины.

Данные, поступающие от каждого источника, имеют определенный шаг изменения значений. Это связано с ограничениями шкалы измерений приборов (последний учитываемый знак), особенностями используемой методики исследования и т.д. Шаг может варьироваться в больших пределах в зависимости от вида исследуемого параметра. Обозначим s_1 шаг измерений

источника D_1 , s_2 – шаг источника D_2 . Шаг измерений можно связать с погрешностями источников: погрешность должна составлять не менее половины шага измерения, т.е. $s_1/2 < \Delta_1$ и $s_2/2 < \Delta_2$.

Пересечение диапазонов $[d_i^1 - \Delta_1; d_i^1 + \Delta_1]$ и $[d_i^2 - \Delta_2; d_i^2 + \Delta_2]$ дает «зону согласия» [5]. Таким образом

$$[d_i^1 - \Delta_1; d_i^1 + \Delta_1] \cap [d_i^2 - \Delta_2; d_i^2 + \Delta_2] = [z_{\min}^i; z_{\max}^i]$$

где $z_{\min}^i$ и $z_{\max}^i$ – нижняя и верхняя границы «зоны согласия» для i -го периода.

При отсутствии ошибок значения обоих источников D_1 и D_2 для i -го периода будут лежать в «зоне согласия». Фактическое значение параметров также будет лежать в этой области. Это позволяет сузить область фактических значений.

В случае, если диапазоны $[d_i^1 - \Delta_1; d_i^1 + \Delta_1]$ и $[d_i^2 - \Delta_2; d_i^2 + \Delta_2]$ не пересекаются, либо значения не входят в «зоны согласия», т.е. $d_i^1 \notin [z_{\min}^i; z_{\max}^i]$ или $d_i^2 \notin [z_{\min}^i; z_{\max}^i]$, можно говорить о наличии ошибок в значениях одного из источников. Определить, какой же из источников содержит значения с ошибками, можно, например, вычислив тренд изменения значений и найдя наибольшее отклонение от него.

При необходимости совместного использования данных, полученных из нескольких источников (например, для заполнения пропусков в периодичности наблюдений), возникает вопрос: как объединить данные? Можно ли просто взять недостающие данные из разных источников и использовать их без предварительной обработки и корректировки? Чаще всего – нет. Пример рис. 1 показал, насколько могут отличаться данные. По этой причине необходима выработка более продуктивного подхода.

Для обеспечения возможностей использования данных из различных источников предлагается использование следующего решения [3, 5]. Вместо того чтобы использовать непосредственное значение показателя, взятого из того или иного источника, мы предлагаем использовать тройку «значение, погрешность, шаг измерения» $\langle d_i^1, \Delta_1, s_1 \rangle$, преобразовывая их при необходимости к единым единицам измерения (например, из г/м³ в мг/дм³). Такое решение позволяет собрать достаточно полную информацию об исследуемых процессах и обеспечить достаточную адекватность данных, собранных из различных источников, фактическому состоянию.

Рассмотрим алгоритм согласования данных мониторинга. В общем виде его можно представить следующим образом:

1. Преобразование данных различных источников к одинаковым единицам измерения.

2. Определение для каждого j -го источника погрешности измерения Δ_j .

3. Определение для каждого j -го источника шага измерения s_j .

4. Преобразование данных в тройку «значение, погрешность, шаг измерения» $\langle d_i^j, \Delta_j, s_j \rangle$.

5. Определение «зоны согласия» источников $[z_{\min}^i, z_{\max}^i]$, где $z_{\min}^i$ – нижняя граница «зоны согласия» для i -го периода, $z_{\max}^i$ – верхняя граница «зоны согласия» для i -го периода.

6. Проверка корректности данных (попадания значений в «зону согласия») $d_i^j \in [z_{\min}^i, z_{\max}^i]$.

7. Корректировка данных источников в случае необходимости (если $d_i^j \notin [z_{\min}^i, z_{\max}^i]$).

8. Приведение данных с разной точностью к единой точности и порядку.

Выявление значимых изменений наблюдаемого параметра

Допустим, что имеется ряд наблюдений параметра за определенный n -й период. Данные наблюдений собраны из разных источников. Возникает вопрос: как выявить значимые изменения параметра в $n + 1$ период при наличии данных из разных источников за предшествующие периоды? Алгоритм выявления значимых изменений наблюдаемого параметра можно представить следующим образом [3].

1. Составление троек «значение, погрешность, шаг измерения» $\langle d_i^j, \Delta_j, s_j \rangle$ для n предшествующих периодов для всех источников.

2. Получение тройки $\langle d_{n+1}^k, \Delta_k, s_k \rangle$ для $n + 1$ периода от источника k .

3. Расчет «зон согласия» источников за n предшествующих периодов.

4. Проверка, существуют ли случаи не-вхождения данных от источника k в «зону согласия», т.е. $d_i^k \notin [z_{\min}^i, z_{\max}^i]$. Если да – источник недостаточно надежен. Требуется дополнительные наблюдения.

5. Проверка, имеются ли тройки $\langle d_n^k, \Delta_k, s_k \rangle, \langle d_{n-1}^k, \Delta_k, s_k \rangle, \dots$ для предшествующих периодов. Если да – переход к п. 6, иначе – п. 7.

6. Если $\frac{|d_{n+1}^k - d_n^k|}{s_k} > 1$ или $|d_{n+1}^k - d_n^k| > \Delta_k$,

то произошли значимые изменения параметра наблюдения.

7. Поиск источника l с минимальным значением $|d_{n+1}^k - d_n^l|$.

8. Если $s_k \geq s_p$, вычисляем $\frac{|d_{n+1}^k - d_n^l|}{s_k}$,

иначе вычисляем $\frac{|d_{n+1}^k - d_n^l|}{s_l}$. Если получен-

ное значение больше 1, то произошли значимые изменения параметра наблюдения.

9. Если при $\Delta_k \geq \Delta_l$ верно $|d_{n+1}^k - d_n^l| > \Delta_k$, или при $\Delta_k < \Delta_l$ верно $|d_{n+1}^k - d_n^l| > \Delta_l$, то произошли значимые изменения параметра наблюдения.

Заключение

Использование тройки «значение, погрешность, шаг измерения» позволяет осуществлять более точное согласование данных. Кроме того, это позволяет совместно использовать данные из различных источников с учетом их специфики. Анализ информации с учетом погрешностей позволяет, с одной стороны, исключить ошибочные выводы о значительных изменениях некоторых параметров наблюдений, а с другой стороны, выявить существенные закономерности в развитии экзогенных процессов. Предлагаемый подход не ограничивается экзогенными процессами и может использоваться в мониторинге других природных процессов и явлений.

Работа выполнена при поддержке гранта РФФИ № 13-07-97510 p_центр_a.

Список литературы

1. Грановский В.А., Сирая Т.Н. Методы обработки экспериментальных данных. – Л.: Энергоатомиздат, 1990.
2. Новицкий П.В., Зograf И.А. Оценка погрешностей результатов измерений. – М.: Энергоатомиздат, 1990.
3. Шарапов Р.В. Алгоритм согласования данных мониторинга экзогенных процессов // Машиностроение и безопасность жизнедеятельности. – 2013. – № 4. – С. 47–49.
4. Шарапов Р.В. Мониторинг экзогенных процессов // Машиностроение и безопасность жизнедеятельности. – 2012. – № 2. – С. 39–42.

5. Шарапов Р.В. О согласовании данных мониторинга экзогенных процессов, полученных из разнородных источников // Машиностроение и безопасность жизнедеятельности. – 2013. – № 4. – С. 43–46.

6. Шарапов Р.В. Переход от технических к природно-техническим системам // Машиностроение и безопасность жизнедеятельности. – 2012. – № 2. – С. 43–46.

7. Шарапов Р.В. Показатели наблюдения и оценки карстовых процессов // Машиностроение и безопасность жизнедеятельности. – 2013. – № 1. – С. 28–34.

8. Sharapov R.V., Kuzichkin O.R. Monitoring of Karst-Suffusion Formation in Area of Nuclear Power Plant // Proceedings of the 7th 2013 IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS). – 12–14 September 2013, Berlin, Germany. – Vol. 2. – 2013. – P. 810–813.

9. Sharapov R.V. The generalized structure of the groundwater monitoring system // 13 international multidisciplinary scientific geoconference SGEM2013. Water resources. Forest, marine and ocean ecosystems. Conference proceedings. 16–22 June 2013, Albena, Bulgaria, 2013. – P. 389–392.

References

1. Granovskiy V.A., Siraya T.N. Metody obrabotki jeksperimental'nyh dannyh [Methods of experimental data processing]. Leningrad, Energoatomizdat, 1990.
2. Novitskij P.V., Zograf I.A. Ocenka pogreshnostej rezul'tatov izmerenij [Estimation of errors of measurement results]. Moscow, Energoatomizdat, 1990.
3. Sharapov R.V. Algoritm soglasovanija dannyh monitoringa jekzogenykh processov [Conformity algorithm for exogenous processes monitoring data] – Mashinostroyeniye i bezopasnost' zhiznedejatel'nosti, 2013, no. 4, pp. 47–49.
4. Sharapov R.V. Monitoring jekzogenykh processov [Monitoring exogenous processes] – Mashinostroyeniye i bezopasnost' zhiznedejatel'nosti, 2012, no. 2, pp. 39–42.
5. Sharapov R.V. O soglasovanii dannyh monitoringa, poluchennyh iz raznorodnyh istochnikov [On the conformity of monitoring data obtained from various sources] – Mashinostroyeniye i bezopasnost' zhiznedejatel'nosti, 2013, no. 4, pp. 43–46.
6. Sharapov R.V. Perehod ot tehicheskikh k prirodno-tehicheskim sistemam [The transition from the technical to the natural-technical systems] – Mashinostroyeniye i bezopasnost' zhiznedejatel'nosti, 2012, no. 2, pp. 43–46.
7. Sharapov R.V. Pokazateli nabljudeniya i ocenki karstovykh processov [Indicators for monitoring and assessment of karst processes] – Mashinostroyeniye i bezopasnost' zhiznedejatel'nosti, 2013, no. 1, pp. 28–34.
8. Sharapov R.V., Kuzichkin O.R. Monitoring of Karst-Suffusion Formation in Area of Nuclear Power Plant // Proceedings of the 7th 2013 IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 12–14 September 2013, Berlin, Germany. Vol. 2, 2013. pp. 810–813.
9. Sharapov R.V. The generalized structure of the groundwater monitoring system // 13 international multidisciplinary scientific geoconference SGEM2013. Water resources. Forest, marine and ocean ecosystems. Conference proceedings. 16–22 June 2013, Albena, Bulgaria, 2013. pp. 389–392.

Рецензенты:

Жизняков А.Л., д.т.н., профессор, первый заместитель директора, Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых, г. Муром;

Орлов А.А., д.т.н., доцент, заведующий кафедрой «Физика и прикладная математика», Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых, г. Муром.

Работа поступила в редакцию 06.03.2014.