

УДК 551.3

О согласовании данных мониторинга экзогенных процессов, полученных из разнородных источников *

Шарапов Р.В.

В работе рассматриваются вопросы обработки данных, полученных из разнородных источников. Такие данные имеют разные форматы, получаются с помощью различных приборов с использованием различных методик. Кроме того, данные могут иметь разную точность и погрешность измерений. Это усложняет их совместное использование и может приводить к существенным ошибкам. По этой причине предлагается вместо использования непосредственных значений показателей, взятых из того или иного источника, применять тройки «значение, погрешность, шаг измерения», преобразовывая их при необходимости к одинаковым единицам измерения. Такое решение позволяет обеспечивать достаточную адекватность данных, собранных из различных источников, фактическому состоянию. Кроме того, учет погрешностей измерений позволяет более точно оценивать динамику процессов. Так, можно произвести фильтрацию колебаний значений того или иного показателя в зоне погрешности и выявить действительно значимые изменения.

Ключевые слова: экзогенные процессы, мониторинг, данные, обработка данных, погрешность, источник.

On the conformity of monitoring data obtained from various sources

Sharapov R.V.

The paper deals with the issues of processing data obtained from different sources. This data formats are different and can be received by means of different devices using various techniques. Furthermore, data may have different accuracy and measurement error. This complicates their sharing and can lead to significant errors. That is why, it is advisable to apply three-item combinations of «value, error, measurement interval», converting and bringing the combinations to common measurement units, if necessary, instead of using direct indicator readings taken from one source or another. This solution enables us to provide sufficient adequacy of data, collected from various sources, to the practical situation. In addition, the measurement error record enables us to assess the processes dynamics more accurately. So, you can filter value instability in the error zone and identify really significant changes.

Keywords: exogenous processes, monitoring, data, data processing, error, information source.

Введение

При проведении мониторинга экзогенных процессов возникает проблема обработки и взаимного использования данных, полученных из различных источников [1].

Анализ показал, что исследования, производимые различными службами в разные годы, имеют существенные различия. Во-первых, перечень оцениваемых параметров отличается среди различных источников информации (служб, проводивших исследования). Во-вторых, перечень параметров

наблюдения для каждой службы может меняться с течением времени (при ретроспективных наблюдениях). В-третьих, различные службы при оценке одних и тех же параметров могут использовать различные методики и оборудование, что приводит к получению результатов с различными погрешностями.

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

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

Проблема согласования данных

В рамках исследования влияния подземных вод на развитие экзогенных процессов [2], в 2013 году были проведены работы по оценке состояния поверхностных и подземных вод г. Муром [3, 4]. В качестве основы для исследований были использованы данные, непосредственно полученные на кафедре Техносферной безопасности, данные Санитарно-гигиенического мониторинга Муромского района, сведения Роспотребнадзора и МУП «Водопровод и канализация». При этом возникли проблемы в связи со сложностями согласования данных, полученных их разных источников.

Рассмотрим проблему на примере параметра «Цветность» воды, который хоть и не является основным с точки зрения экзогенных процессов, в полной мере показывает суть проблемы. Параметр «Цветность» измеряется в градусах и оценивается в различных источниках (использованных нами) с шагом от 1 до 5. Таким образом, если фактическое значение параметра изменяется, например, с 7 до 8 градусов, то при измерениях с шагом 5, итоговое значение увеличится с 5 до 10 (см. рис. 1). Другими словами, будет получена ложная информация об изменении параметра «Цветность» в 2 раза (на 5 градусов), против реального изменения на 1 градус.

Возникает вопрос: информацию какого из источников следует использовать? Конечно, проблема легко может быть решена путем выбора наиболее представительного источника (обеспечивающего высокую точность результатов). Но что делать, если не один из источников не предоставляет данные за весь анализируемый период (например, часть данных отсутствует или периодичность их получения больше интервала мониторинга) или наиболее информативный источник (предоставляющий наиболее полную информацию) имеет наименьшую точность?

Согласование данных

Рассмотрим рис. 1. Пусть имеется два источника 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 1 за i -й период, d_i^2 – данные наблюдений Источника 2 за i -й период. Данные каждого источника за i -й период не совпадают друг с другом [5, 6].

Рассмотрим погрешности измерений каждого источника. Обозначим погрешность измерений Источника 1 как Δ_1 , погрешность Источника 2 как Δ_2 [7, 8]. Тогда фактическое значение параметров от Источника 1 будет лежать в диапазоне $[d_i^1 - \Delta_1; d_i^1 + \Delta_1]$ («зоне погрешности»), от Источника 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.

Каждый источник представляет данные с определенным шагом измерений. Обозначим s_1 шаг измерений Источника 1, s_2 – Источника 2. Конечно, шаг измерений можно связать с погрешностями источников: погрешность должна составлять не менее половины шага измерения, т.е. $s_1/2 < \Delta_1$ и $s_2/2 < \Delta_2$.

Как можно заметить, значения обоих источников для i -го периода будут лежать в зоне пересечения погрешностей («зоне согласия»). Кроме того, фактическое значение параметров также будет лежать в этой области. Это позволяет сузить область фактических значений. Кроме того, подобный подход позволяет выявить ошибки измерений. Так, непопадание значения какого-то из источников в «зону согласия» или не пересечение зон погрешностей для i -го периода свидетельствует об ошибках в одном из источников.

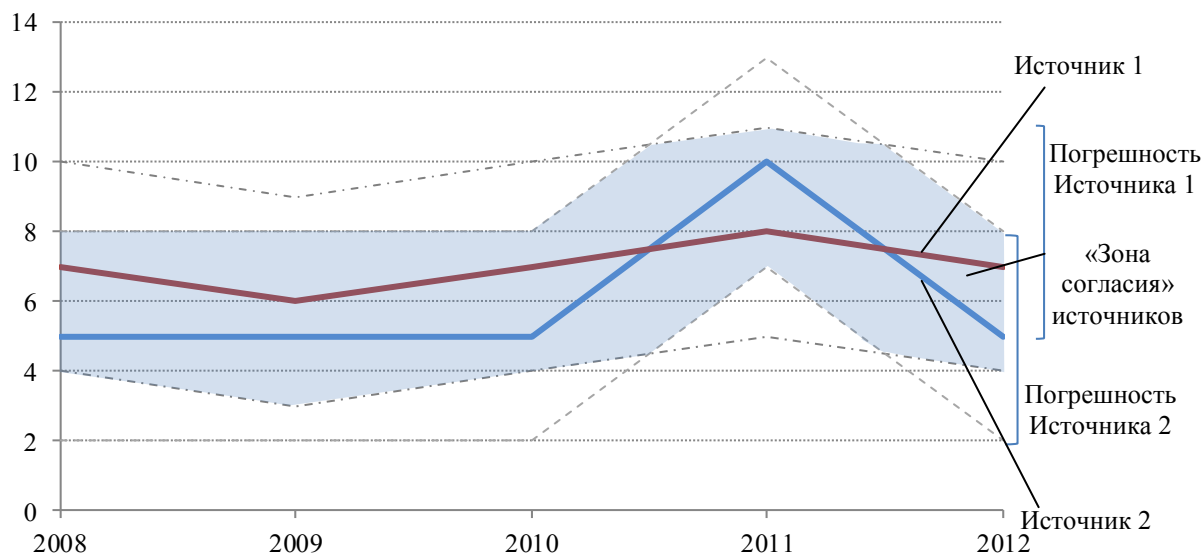


Рис. 1. Согласование данных разных источников информации.

Рассмотрим теперь вопрос выбора базового источника информации при отсутствии полных данных за весь анализируемый период. Наиболее приемлемым решением является выбор источника, обеспечивающего достаточно высокую точность и полноту данных, и дополнение его данными из других источников. Тем не менее, при этом также возникает проблема: следует ли проводить преобразования и согласования данных, полученных из разных источников? Несмотря на сложность и неформализованность таких преобразований, очевидно, что их реализация необходима. В противном случае результаты могут быть значительно искажены.

Для обеспечения возможностей использования данных из различных источников предлагается использование следующего решения. 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) [5, 6]. Такое решение позволяет собрать достаточно полную информацию об исследуемых процессах и обеспечить достаточную адекват-

ность данных, собранных из различных источников, фактическому состоянию.

Заключение

Таким образом, 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.

Литература

1. Шаронов Р.В. Мониторинг экзогенных процессов // Машиностроение и безопасность жизнедеятельности, 2012, № 2. – С.39-42
2. Шаронов Р.В. Переход от технических к природно-техническим системам // Машиностроение и безопасность жизнедеятельности, 2012, № 2. – С.43-46.
3. Димакова Н.А., Шаронов Р.В. Проблема загрязнения подземных вод // Современные наукоемкие технологии, 2013, № 2. – С. 79-82.

4. Шарпов Р.В. Принципы мониторинга подземных вод // Машиностроение и безопасность жизнедеятельности, 2012, № 3. – С.27-30.

5. Шарпов Р.В. Структура системы мониторинга подземных вод // Машиностроение и безопасность жизнедеятельности, 2012, № 4. – С.20-23.

6. 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.

7. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. – М.: Энергоатомиздат, 1990.

8. Грановский В.А., Сирая Т.Н. Методы обработки экспериментальных данных. – Л.: Энергоатомиздат, 1990.

References

1. Sharapov R.V. Monitoring exogenous processes // Engineering industry and life safety, 2012, № 2. – P.39-42.

2. Sharapov R.V. The transition from the technical to the natural-technical systems // Engineering industry and life safety, 2012, № 2. – P.43-46.

3. Dimakova N.A., Sharapov R.V. The problem of groundwater pollution // Modern high technologies, 2013, № 2. – P. 79-82.

4. Sharapov R.V. Principles of groundwater monitoring // Engineering industry and life safety, 2012, № 3. – P.27-30.

5. Sharapov R.V. The structure of the groundwater monitoring system // Engineering industry and life safety, 2012, № 4. – P.20-23.

6. 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.

7. Novitskiy P.V., Zograph I.A. Estimation of errors of measurement results. – Moscow: Energoatomizdat, 1990.

8. Granovskiy V.A., Siraya T.N. Methods of experimental data processing. – Leningrad: Energoatomizdat, 1990.

Статья поступила в редакцию 19 октября 2013 г.

Шарпов Руслан Владимирович – кандидат технических наук, заведующий кафедрой «Техносферная безопасность» Муромского института (филиала) федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых», г. Муром, Россия. E-mail: info@vanta.ru

Sharapov Ruslan Vladimirovich – Ph.D., Murom Institute of Vladimir State University, Murom, Russia. E-mail: info@vanta.ru