

Раздел I. Моделирование систем и процессов

УДК 519.24

О.А. Попова

ГИСТОГРАММЫ ВТОРОГО ПОРЯДКА ДЛЯ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ В ЗАДАЧАХ С ИНФОРМАЦИОННОЙ НЕОПРЕДЕЛЕННОСТЬЮ

Решение многих практических задач в условиях информационной неопределенности входных данных требует специальных методов, основанных на процедурах представления и численного моделирования. Рассматривается процедура распространения неопределенности (propagation of uncertainty) и приводится анализ существующих методов ее представления. Для решения таких задач предлагается использовать численный вероятностный анализ. Численный вероятностный анализ представляет собой способ распространения информационной неопределенности, в том числе для задач, когда вероятностные оценки входных параметров носят неопределенный характер. С целью снижения уровня информационной неопределенности и получения дополнительной информации о распределении параметров в условиях информационной недостаточности, предлагается использовать гистограммный подход. Представление неопределенности, содержащейся в параметрах входных данных, осуществляется с использованием гистограмм второго порядка, на основе которых строятся процедуры ее распространения. С этой целью на основе гистограмм второго порядка разработана арифметика неопределенных данных. Приводятся численные примеры и обсуждается практика применения.

Информационная неопределенность; процедуры распространения неопределенности; численный вероятностный анализ; гистограммы второго порядка.

О.А. Popova

SECOND ORDER HISTOGRAM FOR NUMERICAL SIMULATION PROBLEMS WITH INFORMATION UNCERTAINTY

The solution of many practical problems with information uncertainty of input data requires special techniques based on the submission procedures and numerical simulation. The article describes the uncertainty propagation procedures (propagation of uncertainty) and an analysis of existing methods of its representation. To solve such problems are encouraged to use numerical probabilistic analysis. Numerical probabilistic analysis is a way for propagation of information uncertainty, including problems when probabilistic estimates of the input parameters are uncertain. To reduce the level of information uncertainty and to have more information about the distribution of the parameters in an information insufficiency is proposed to use the histogram approach. Representation of uncertainty contained in the input data is performed using a second-order histograms, which are constructed on the basis of its distribution procedures. For this purpose, based on a second order histogram is developed the arithmetic of undefined data. There are numerical examples and discussing the practical applications.

Information uncertainty; uncertainty propagation procedures; numerical probabilistic analysis; second order histogram.

Анализ существующих подходов к представлению и распространению информационной неопределенности. Обработка информации и анализ систем в условиях информационной неопределенности требует применения специальных

методов исследования. В настоящее время во многих работах, посвященных исследованию систем в условиях информационной неопределенности, используется подход, который в зарубежных источниках получил название «propagation of uncertainty» [1–3]. В дословном переводе на русский язык это звучит как «распространение неопределенности». Анализируя содержательный смысл этого понятия, можно выделить важный аспект процедуры распространения неопределенности, как способ получения дополнительных оснований (знаний) для исследования входных данных. Другими словами, если исследователь находится в условиях недостаточности или полного отсутствия эмпирической информации или оснований для выдвижения идей и предположений о неизвестном распределении входных параметров, то необходимо распространить (propagate) существующую неопределенность, чтобы получить достаточные выводы в соответствии с принципом недостаточного основания. Чтобы получить необходимые основания для оценки или восстановления неизвестного входного распределения на основе неполной, неточной информации, можно использовать различные процедуры, например, следует рассмотреть распределение вероятностей, которое имеет максимальную энтропию, допускаемую имеющейся априорной информацией. При этом важно следовать принципу: выбрать тип представления в соответствии с количеством имеющейся информации и оставаться верным имеющейся информации, включая информационные пробы. Применение принципа недостаточности (достаточности) оснований позволяет существенно расширить формы представления неопределенностей. Например, р-боксы [1], облака [2], теория Демпстера-Шафера [3], интервальные гистограммы, гистограммы второго порядка [4].

В рамках рассмотренных принципов “распространение неопределенности” можно достичь, используя метод вероятностных границ (Probability bounds). Его основная идея приводится, например, в работе [1], и состоит в следующем: “Есть нечто, что можно сказать о неизвестном распределении. В частности, его кумулятивная функция распределения вероятностей или CDF (Cumulative Distribution Function) должна лежать в области – ящике (box), ограниченная нулем и единицей по вертикали и от минимума и максимума горизонтально. Истинная функция распределения, какой бы она ни была, должна находиться в этой области”. Идея построения вероятностных границ оказалась весьма продуктивной и нашла свое применение в такой форме представления неопределенности, как р-боксы (р-box Фэрсона Скотта). Еще одним подходом к распространению неопределенности являются Облака Неймайера (Neumaier’s clouds), которые позволяют представить неполную стохастическую информацию четким, понятным и вычислительно привлекательным способом, а также позволяют визуализировать неопределенность и обладают четкой семантикой, выступая посредником между понятием нечеткого множества и вероятностным распределением. В рамках основных подходов к распространению неопределенности следует указать также на математическую теорию очевидностей (свидетельств) Демпстера-Шафера, основанную на функции доверия (belief functions) и функции правдоподобия (plausible reasoning), которые используются, чтобы скомбинировать отдельные части информации (свидетельства) для вычисления вероятности события. Данная теория позволяет построить необходимые основания в условиях неопределенности, путем оценки верхней и нижней границы интервала возможностей. Среди подходов к распространению неопределенностей следует особенно выделить метод, который опирается на понятие «вероятность второго порядка», и известен как second-order probability. Данный подход представляет собой метод, позволяющий строить вероятностные оценки в случае эпистемистической неопределенности. Концепция вероятностей

второго порядка была изложена в 1996 г. в работах А. Mosleh и V.M. Bier. Анализ публикаций показал, что, несмотря на то, что данное направление достаточно активно развивается за рубежом, понятие «вероятность второго порядка» еще находится в стадии определения [5].

Распространение неопределенности на основе численного вероятностного анализа. Численный вероятностный анализ (ЧВА) представляет собой еще один способ распространения информационной неопределенности, в том числе для задач, когда вероятностные оценки входных параметров носят неопределенный характер. С целью снижения уровня информационной неопределенности и получения дополнительной информации о распределении параметров в условиях информационной недостаточности предлагается использовать гистограммный подход.

Идея гистограммного подхода изложена в работах [4, 6, 7], и заключается в следующем: наряду с общими представлениями случайных величин своими плотностями в виде непрерывных функций, можно рассматривать случайные величины, плотность распределения которых представляет гистограмму.

Для решения таких задач можно также использовать интервальные гистограммы и гистограммы второго порядка. В тех случаях, когда нет возможности получить точную функцию распределения случайной величины, задают оценки плотности распределения сверху и снизу. Такие оценки удобно аппроксимировать *интервальными гистограммами* [4]. Гистограмму будем называть интервальной, если значения гистограммы принимают интервальные значения.

Наряду с интервальными гистограммами в рамках ЧВА предлагается использовать *гистограммы второго порядка*, т.е. такие гистограммы, каждый столбец которых – гистограмма [4]. Определим гистограмму второго порядка (ГВП) как кусочно-гистограммную функцию. ГВП так же, как и обычная гистограмма определяется сеткой $\{z_i, i=0,1,\dots,n\}$ и набором гистограмм $\{P_i, i=1,2,\dots,n\}$. На каждом отрезке $[z_{i-1}, z_i]$ ГВП принимает гистограммное значение P_i .

Рассмотрим процедуру построения ГВП. Пусть мы имеем ряд гистограмм $\{Y_i, i=1,2,\dots,N\}$. Каждой Y_i поставим в соответствие вероятность p_i : $\sum p_i=1$. Для простоты будем считать, что все гистограммы Y_i заданы на сетке $\{z_i, i=0,1,\dots,n\}$ и на отрезке $[z_{k-1}, z_k]$ Y_i принимает значение Y_{ik} . Таким образом, на каждом отрезке $[z_{k-1}, z_k]$ имеем случайную величину Y_k , принимающую значения Y_{ik} с вероятностью p_i . Используя эти значения, мы можем на каждом отрезке $[z_{k-1}, z_k]$ восстановить гистограмму P_{z_k} .

Рассмотрим пример задачи, когда вероятностные характеристики изучаемого показателя содержат неопределенности, и построим для него гистограмму второго порядка.

Пример 1. Пусть параметр S представляет собой некоторую случайную величину S_t , которая имеет треугольное распределение P_t на отрезке $[0, 1]$, высота $h = 2$ и вершина в некоторой точке, $(t, 2)$, где t – случайная величина с треугольным распределением на отрезке $[0,25; 0,75]$ с вершиной $(0,5; 4)$ (рис. 1,а).

На рис. 1,а представлена случайная величина S_t , на рис. 1,б приведена гистограмма второго порядка, где оттенками серого показано распределение вероятностей. Интервальное распределение (максимальное и минимальное P_t для всех t) изображено граничными линиями. Внутренняя линия определяет «эффективную» плотность вероятности гистограммы второго порядка – математическое ожидание плотностей вероятности P_t в точке x .

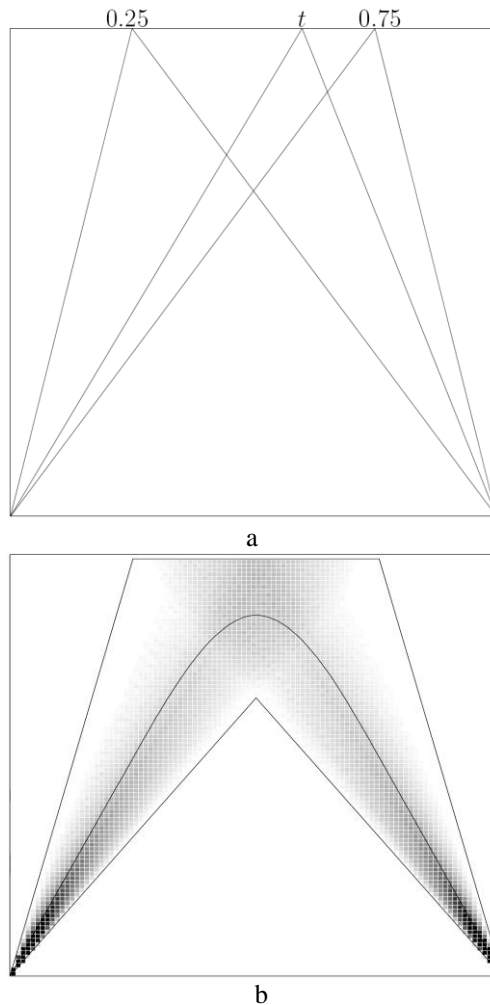


Рис. 1. Гистограмма второго порядка для параметра S

Арифметика неопределенных данных на основе ГВП. Для осуществления численных операций над «неопределенными» переменными, заданными своими функциями плотности в виде гистограмм второго порядка, в условиях неопределенности, определим арифметику для ГВП.

Пусть X, Y – ГВП, определяются сетками $\{v_i, i=0,1,\dots,n\}$, $\{w_i, i=0,1,\dots,n\}$ и наборами гистограмм $\{P_{x_i}\}$, $\{P_{y_i}\}$. Пусть $Z=X*Y$, где $*$ $\in \{+, -, \cdot, /, \hat{\uparrow}\}$. Построим Z как ГВП. Зададим сетку $\{z_i, i=0,1,\dots,n\}$, тогда гистограмма P_{z_i} на отрезке $[z_{i-1}, z_i]$, следуя работе [1], определяется по формуле

$$P_{z_i} = \int_{\Omega_i} X(\xi)Y(\eta)d\xi d\eta,$$

где $\Omega_i = \{(\xi, \eta) \mid z_i \leq \xi * \eta \leq z_{i+1}\}$. Заметим, что на каждом прямоугольнике $[v_{i-1}, v_i] \times [w_{j-1}, w_j]$ функция $X(\xi)Y(\eta)$ – есть постоянная гистограмма $P_{x_i}P_{y_j}$. Интеграл от гистограммы по некоторой области – есть значение гистограммы, умноженное на площадь области.

Проиллюстрируем, как работает гистограммная арифметика в случае сложения двух ГВП.

Пример 2. Пусть необходимо сложить две гистограммы второго порядка X и Y . Гистограммы X и Y порождены равномерными случайными величинами, заданными соответственно на отрезках $[0, t_1]$ и $[t_2, 2]$, где t_1 – равномерная случайная величина, заданная на отрезке $[1, 2]$, t_2 – равномерная случайная величина, заданная на отрезке $[0, 1]$. Результат сложения двух гистограмм представлен в виде гистограммы второго порядка Z , изображенной на рис. 2. Носителем Z является отрезок $[0, 4]$, высота 1, значения плотности вероятности представлены оттенками серого.

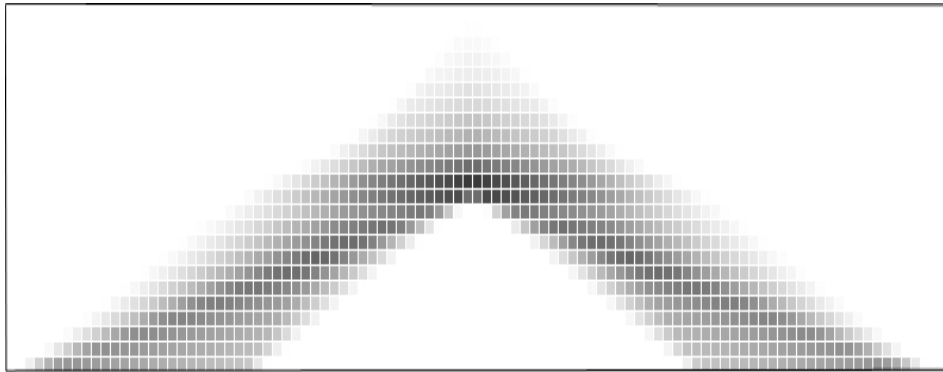


Рис. 2. Сумма двух гистограмм второго порядка $Z = X+Y$

Использование гистограмм второго порядка в условиях неопределенности вероятностных характеристик параметров системы имеет широкую область применения к решению разнообразных практических задач. Например, к решению задач оценки эффективности инвестиционных проектов. В данном случае мы имеем дело с высоким уровнем рыночной неопределенности, где стандартная финансовая модель не учитывает вероятностную природу характеристик оцениваемых показателей и соответственно не может дать достаточных оснований для принятия эффективных инвестиционных решений, а используемые методы, например, метод дисконтированных денежных потоков, не учитывает вероятностный характер результатов инвестиционных проектов. Гистограммы второго порядка также могут успешно применяться в задачах оценки показателей надежности и оценки безотказной работы сложных технических систем, для изучения гидрологических и других систем. Решение практических задач с использованием методов интерполяции и экстраполяции также лежит в сфере применения ГВП. Рассмотрим примеры, где возникает необходимость использования ГВП.

Задачи интерполяции. При численном моделировании часто приходится пользоваться различными зависимостями между входными и выходными величинами – в том числе как эмпирическими зависимостями, построенными на основе наблюдений, так и получающимися в процессе исследования. Для этого широко применяются методы интерполяции и экстраполяции. Например, задачи интерполяции возникают при подборе эмпирической формулы по данным измерения. Здесь надо предостеречь от формального, слепого подбора такой формулы только по измеренным значениям. Выбор вида формулы (многочлен, степенная функция, экспонента и т.д.) должен опираться не только на теоретическое обсуждение различных свойств изучаемой зависимости, но и на анализ имеющейся информации с точки зрения ее полноты и специфических особенностей присутствующей в ней неопределенности. Это значит проанализировать множество аспектов, связанных как с понятием “неоп-

ределенная информация”, так и собственно самими численными методами, предназначенными для ее представления, обработки, преобразования и извлечения знаний. С этой точки зрения, следует выделить два важных аспекта, которые следует иметь в виду исследователю при работе с неопределенными данными. Первый аспект связан с понятием “неопределенная информация” и проблемой изучения источников неопределенности и снижения уровня неопределенности в данных. Вторым аспектом связан с выбором численного метода для реализации вычислительных процедур, которые могли бы существенно помочь исследованию, а не оказаться источником ошибок. В рамках поставленных проблем рассмотрим применение ЧВА к задачам интерполяции и экстраполяции в условиях стохастической неопределенности.

Рассмотрим случай линейной интерполяции. Пусть для некоторой функции f в точках x_1, x_2 значения f_1, f_2 известны не точно и представлены независимыми гистограммами F_1 и F_2 .

В случае линейной интерполяции имеем точное равенство

$$f(x) = l_1(x) + \frac{(x-x_1)(x-x_2)}{2} f''(\xi),$$

где l_1 – полином Лагранжа первой степени, $\xi \in [x_1, x_2]$ и

$$l_1(x) = f_1 \frac{x_2 - x}{x_2 - x_1} + f_2 \frac{x - x_1}{x_2 - x_1}.$$

Необходимо построить линейную гистограммную функцию $L(x)$ такую, что выполнены условия интерполяции $L(x_1) = F_1$ и $L(x_2) = F_2$. Таким образом, используя естественные вероятностные расширения, построим гистограммный полином Лагранжа первой степени

$$L(x) = F_1 \frac{x_2 - x}{x_2 - x_1} + F_2 \frac{x - x_1}{x_2 - x_1}.$$

Заметим, что условия интерполяции выполнены и $L(x)$ принимает соответствующие значения в узлах интерполяции.

Далее, если необходимо построить гистограммную функцию f , для которой выполнено включение $f \in \mathbf{f}, \forall x \in [x_1, x_2]$, то необходимо знать априорную информацию о плотностях вероятности f'' на отрезке $[x_1, x_2]$. Заметим, что

$$f(x) = L(x) + \frac{(x-x_1)(x-x_2)}{2} f''(\xi).$$

Поскольку точка ξ не известна, то для оценки f'' воспользуемся гистограммами второго порядка. Для этих целей построим гистограмму второго порядка $\mathbf{f}'' = \{f''(x) | x \in [x_1, x_2]\}$.

Следуя гистограммному подходу для плотности вероятности $f(x)$, получаем включение

$$f(x) \in L(x) + \frac{(x-x_1)(x-x_2)}{2} \mathbf{f}''.$$

Таким образом, построение включений с использованием простейших интерполяционных формул приводит к необходимости расширить стандартные действия над гистограммными переменными элементами гистограмм второго порядка и использовать процедуру распространения неопределенности. Последнее не сильно усложняет работу гистограммного анализа.

Практика применения ГВП в экономике. Рассмотрим простой пример, когда некоторая фирма собирается производить новый продукт и необходимо оценить денежные потоки от его реализации. Пусть планируется продать $\bar{N} = 1000$ единиц продукции по цене $\bar{C} = 10$ условных единиц и получить от реализации продукта соответственно выручку $R = 10000$ условных единиц. Однако в силу рыночной неопределенности N , C – случайные величины. Допустим, экспертные оценки при этом показывают, что величина N будет иметь закон распределения

$$P_N = 2\left(\sum_{i=1}^n t_i / n - 0.5\right)\Delta + \bar{N},$$

с математическим ожиданием $\bar{N} \in [950, 1050]$, с носителем $[\bar{N} - \Delta, \bar{N} + \Delta]$, $\Delta = 50$, где t_i – равномерно распределенные случайные величины на отрезке $[0, 1]$, n – случайное целое в диапазоне от 3 до 6. Цена C при этом коррелирована с величиной продаж N следующим образом:

$$C = (N - \bar{N})/\Delta + \bar{C}.$$

В данной задаче относительно показателей N и C , имеющих стохастическую природу, присутствует интервальная неопределенность, которая порождается \bar{N} , что существенно ограничивает информационную составляющую для принятия управленческого решения. Для повышения эффективности управленческого решения применим процедуру распространения неопределенности. Для этого используем гистограммный подход и ГВП – арифметику. Для этого предположим, в силу принципа максимальной энтропии [1], что \bar{N} имеет равномерный закон распределения.

На рис. 3 приведены результаты расчетов, представленные в виде ГВП. По горизонтали – значения R , по вертикали – значения плотности вероятности. Если применить интервальный анализ, то получим, что значения R отличны от нуля на отрезке $[8100, 12100]$. Гистограмму второго порядка определим сеткой $\{z_i = 8100 + hi, i = 0, 1, \dots, 100; h = 40\}$ и набором гистограмм $\{P_i, i = 1, 2, \dots, 100\}$. На каждом отрезке $[z_{i-1}, z_i]$ ГВП принимает гистограммное значение P_i , которое на рис. 3 представлено оттенками серого цвета. Ниже на рис. 4 приведен 50-й столбец гистограммы второго порядка P_{50} .

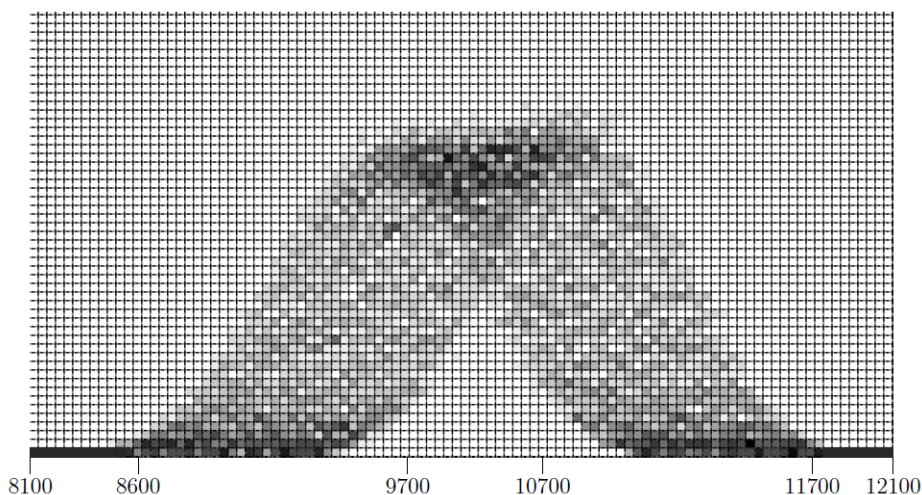
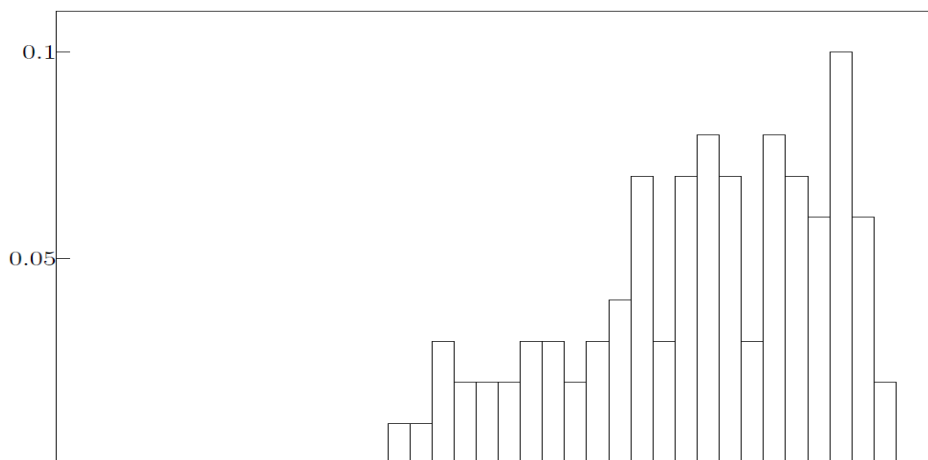


Рис. 3. Гистограмма второго порядка R

Рис. 4. Гистограмма P_{50}

Из анализа гистограммы второго порядка для R следует, что наиболее вероятные значения R лежат в интервале $[9700, 10700]$, значения R из интервалов $[8100, 8600]$ и $[11700, 12100]$ имеют практически нулевые вероятности.

Заключение. Результаты проведенных исследований и численные примеры показывают возможность использования гистограмм второго порядка для представления различных типов информационной неопределенности и позволяют проводить численное моделирование на основе процедур распространения неопределенности для построения дополнительных оснований с целью снижения уровня неопределенности, присутствующей в исходной информации, в частности, для задач с неопределенными параметрами вероятностных характеристик входных данных и задач стохастической интерполяции.

Численные примеры показали эффективность применения ГВП и численного вероятностного анализа для исследования сложных систем в условиях неопределенных данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Ferson S., Ginzburg L.* Different methods are needed to propagate ignorance and Variability // *Reliability Engineering and System Safety*. – 1996. – № 54. – С. 133-144.
2. *Neumaier A.* Clouds, Fuzzy Sets and Probability Intervals // *Reliable Computing*. – 2004. – № 10. – С. 249-272.
3. *Dempster A.P.* Upper and lower probabilities induced by a multi-valued mapping // *Annals of Mathematical Statistics*. – 1967. – № 38. – С. 325-339.
4. *Dobronets B.S., Krantsevich A.M., Krantsevich N.M.* Software implementation of numerical operations on random variables // *Journal of Siberian Federal University. Mathematics & Physics*. – 2013. – № 6 (2). – С. 168-173.
5. *Skyrms B.* Higher Order Degrees of Belief // *Prospects for Pragmatism: Essays in Memory of F.P. Ramsey, D.H. Mellor, ed.* Cambridge; New York: Cambridge University Press. – 1980. – P. 109-137.
6. *Попова О.А.* Технология извлечения и визуализации знаний на основе численного вероятностного анализа неопределенных данных // *Информатизация и связь*, – 2013. – № 2. – С. 63-66.
7. *Popova O.A.* Optimization problems with random data // *Journal of Siberian Federal University. Mathematics & Physics*. – 2013. – № 6 (4). – С. 506-515.

Статью рекомендовал к публикации д.т.н., профессор Ю.И. Rogozov.

Попова Ольга Аркадьевна – Саяно-Шушенский филиал Федерального государственного автономного образовательного учреждения высшего профессионального образования «Сибирский Федеральный университет»; e-mail: olgaarc@yandex.ru; 660041, г. Красноярск, Свободный, 79; тел.: 89135985288; кафедра фундаментальной подготовки; к.т.н.; доцент.

Popova Olga Arkadevna – Sayano-Shushunsky Branch of Federal State-owner Autonomy Educational Establishment of Higher Vocational Educational “Siberian Federal University”; e-mail: olgaarc@yandex.ru; 79, Svobodny, Krasnoyarsk, 660041, Russia; phone: +79135985288; the department of basic training; cand. of eng. sc.; associate professor.

УДК 519.24

Б.С. Добронез, О.А. Попова

ГИСТОГРАММНЫЙ ПОДХОД К ПРЕДСТАВЛЕНИЮ И ОБРАБОТКЕ ДАННЫХ КОСМИЧЕСКОГО И НАЗЕМНОГО МОНИТОРИНГА

Рассматривается проблема изучения природных процессов на основе данных космического и наземного мониторинга. Процесс обработки спутниковых данных для исследования природных процессов включает в себя ряд вычислительных процедур, которые должны отвечать системе требований, среди которых в первую очередь необходимо отметить снижение уровня неопределенности в данных, достоверность и наглядность полученных результатов. На основе численного вероятностного анализа предлагается концептуально – гистограммный подход, который применяется для разработки процедур представления и обработки информационных потоков, а также для численного моделирования и представления характеристик природных объектов. Показывается, что применение разработанных процедур позволяет агрегировать данные, снижает уровень информационной неопределенности в данных и существенно повышает эффективность численных расчетов. Численный вероятностный анализ представляет собой непараметрический подход и может успешно применяться для вероятностного описания систем в рамках визуально-интерактивного моделирования, повышая тем самым качество исследования систем. На тестовых примерах и ряде практических задач доказаны преимущества данного подхода перед методом Монте–Карло.

Численный вероятностный анализ; космический мониторинг; моделирование природных явлений; гистограммный подход; обработка данных.

B.S. Dobronets, O.A. Popova

HISTOGRAM APPROACH TO REPRESENTATION AND PROCESSING OF DATA SPACE AND DATA GROUND MONITORING

We study the natural processes on the basis of space and ground-based monitoring. Processing of satellite data for the study of natural processes involves a number of computational procedures to meet system requirements, such as reducing the level of uncertainty in the data, the accuracy and clarity of the results. On the base of Numerical probabilistic analysis is proposed conceptually – histogram approach, which is used to develop procedures for the representation and for processing of information flows, as well as for numerical modeling and representation of the output characteristics of natural objects. It is shown that the application of the procedures developed allows to aggregate the data, reduces the level of information uncertainty and significantly improves the efficiency of the numerical calculations. Numerical probabilistic analysis is a nonparametric approach can be successfully applied to a probabilistic description of systems within a visual interactive simulation, thereby increasing the quality of the research systems. By test examples and some practical problems proved the advantages of this approach over Monte Carlo.

Numerical probabilistic analysis; modeling of the natural processes; space monitoring; histogram approach; data processing.