

5. Назимова Д. И., Коротков И. А., Чередникова Ю. С. Основные высотно-поясные подразделения лесного покрова в горах Южной Сибири и их диагностические признаки // Структура и функционирование лесных биогеоценозов Сибири. — М.: Наука, 1987.
6. Сочава В. Б. Классификация растительности как иерархия динамических систем // Геоботаническое картографирование. — Л.: Наука, 1978.
7. Зоны и типы поясности растительности России и сопредельных территорий: Карта м-ба 1:8 000 000 и пояснительный текст / Под ред. Г. Н. Огуревой, Т. В. Котовой. — М.: Центр «Интеграция», 1999.
8. Назимова Д. И., Ноженкова Л. Ф., Поликарпов Н. П. Биоклиматические модели и их применение для прогноза трансформаций лесного покрова Сибири. — Красноярск, 1998.
9. Шумилова Л. В. Ботаническая география Сибири. — Томск: Изд-во Том. ун-та, 1962.
10. Ермаков Н. Б. Разнообразие бореальной растительности Северной Азии (континентальные гемибореальные леса, классификация и ординация). — Новосибирск, 2003.
11. Коновалова М. Е. Восстановительно-возрастная динамика низкогорных лесов приенисейской части Восточного Саяна: Автореф. дис. ... канд. биол. наук. — Красноярск, 2004.
12. Колесников Б. П. Геоботаническое районирование Дальнего Востока и закономерности размещения его растительных ресурсов // Вопросы географии Дальнего Востока. — Хабаровск, 1963. — Сб. 6.
13. Смагин В. Н. Лесорастительное районирование Сибири. — Красноярск: Изд-во Ин-та леса, 1977.
14. Исаченко А. Г., Шляпников А. А., Робозерова О. Д. и др. Ландшафтная карта СССР м-ба 1:4 000 000. — М.: ГУГК, 1988.
15. Исаченко А. Г., Шляпников А. А. Ландшафты. — М.: Мысль, 1989.
16. Holdridge L. R. Determination of world plant formations from simple climatic data. — Washington, 1947. — Vol. 105.
17. Федоров В. Д., Гильманов Т. Г. Экология. — М.: Изд-во Моск. ун-та, 1980.
18. Куминова А. В., Седельников В. П., Маскаев Ю. Н. и др. Растительный покров и естественные кормовые угодья Тувинской АССР. — Новосибирск: Наука, 1985.
19. Белов А. В., Лямкин В. Ф., Соколова Л. П. Картографическое изучение биоты. — Иркутск, 2002.
20. Мальшев Л. И., Пешкова Г. А. Особенности и генезис флоры Сибири (Предбайкалье и Забайкалье). — Новосибирск: Наука, 1984.

Институт леса СО РАН,
Красноярск

Поступила в редакцию
1 февраля 2005 г.

УДК 910.2

А. В. ИГНАТОВ

ИНФОРМАЦИОННО-ВЕРОЯТНОСТНОЕ МОДЕЛИРОВАНИЕ В ГЕОГРАФИЧЕСКИХ ИССЛЕДОВАНИЯХ

Дано определение информационно-вероятностного моделирования и приведен перечень основных его этапов. Рассмотрены содержание этапов и их особенности, касающиеся построения и использования информационно-вероятностных моделей. Рекомендованы способы решения типичных задач, возникающих в процессе моделирования.

The definition of the information-probability simulation, and a list of its basic stages are presented. The content of each stage is considered, as well as their salient features concerning the construction and employment of information-probability models. Methods of solving typical problems arising in the simulation process are suggested.

Информационно-вероятностное моделирование — это отображение реальной информации об изменчивости и взаимосвязи характеристик объекта в виде оценки их совместного распределения вероятностей и ее использование для нахождения требуемых условных значений переменных модели. При решении географических задач достоинства такого подхода проявляются в том, что он дает возможность интегрировать в моделях самые различные сведения об изменчивости и взаимосвязи параметров изучаемых объектов и позволяет сравнительно легко и алгоритмически единообразно решать задачи применения моделей путем вычисления соответствующих условных характеристик на основе оценки их совместного распределения. Он способствует более осторожной и правильной

интерпретации результатов моделирования, поскольку требует обязательного учета точности и достоверности исходных данных и, соответственно, моделей и модельных оценок. Предлагаемый подход не исключает использования других моделей (математических, имитационных, графических, натуральных и т. д.) как специфических оценок, содержащих информацию о некоторых характеристиках совместного распределения вероятностей контролируемых переменных.

Процесс информационно-вероятностного моделирования, как и математического или имитационного, можно разделить на основные этапы (см. рис. 1). Обсудим содержание этих этапов и подчеркнем их особенности, характерные для данного случая.

Формулировка цели моделирования. На этом этапе определяется приоритетная функция (описательная, предсказательная или эвристическая), для выполнения которой разрабатывается модель. Если приоритетная функция описательная, то цель моделирования — систематизация данных и составление наиболее наглядного и понятного для потребителя представления об объекте. В географии наиболее типичной моделью в этом случае является карта. Из приоритета предсказательной функции вытекают первоочередность выбора прогнозируемых переменных и их вероятных предикторов, требования к точности и надежности прогноза.

В географических исследованиях данные обстоятельства чаще всего проявляются при разработке моделей экстраполяции временных рядов [1, 2]. При решении эвристических задач проверки гипотез или изучении модели как самостоятельного объекта технологию ее построения следует ориентировать на доказательность получаемых результатов. Среди моделей такого типа наиболее характерны математические, представляющие собой совместную логически упорядоченную запись (систему) ряда условий (уравнений, неравенств и т. д.), накладываемых на значения контролируемых переменных [3].

Формулировка цели определяет рамочные условия для выбора необходимых исходных данных, методов моделирования и форм представления его результатов. Цель задает критерии, с помощью которых можно выявить отношение к предмету исследования используемой для построения модели информации и ее ценность [4].

Накопление необходимых сведений о моделируемом объекте. Источниками информации в этой задаче могут служить сам объект и все его существующие модели. Критерии отбора необходимых данных определяются в соответствии с целью моделирования. Процедура сбора информации состоит в обеспечении ее доступности для разработчика модели. Ее результатом обычно является набор объединенных по теме, но различных по форме и содержанию сведений об объекте. При информационно-вероятностном подходе обязательна оценка точности и достоверности таких данных.

Можно выделить два типовых подхода к обеспечению полноты сведений, необходимых для достижения выбранных целей. Один из них состоит в разработке модели, для которой достаточно имеющейся информации. При другом информационные пробелы, выявляющиеся по мере разработки модели, заполняются специально собираемыми сведениями.

Недостаток первого подхода заключается в том, что разработанная модель может оказаться недостаточно информативной для достижения поставленных целей, если при сборе информации были пропущены важные ее составляющие, ценность которых априорно не представлялась значительной.

Недостаток второго пути обусловлен тем, что при его реализации определяющим фактором информационного обеспечения модели становится не цель работы, а используемая технология моделирования. В этом случае неудачный ее выбор или ошибки в ее выполнении приводят к неоправданным затратам на сбор ненужной информации. Кроме того, сам процесс моделирования становится неустойчивым, так как требуемая по мере его осуществления новая информация может быть не найдена, что вызывает необходимость ее компенсации гипотетическими или случайными данными, или

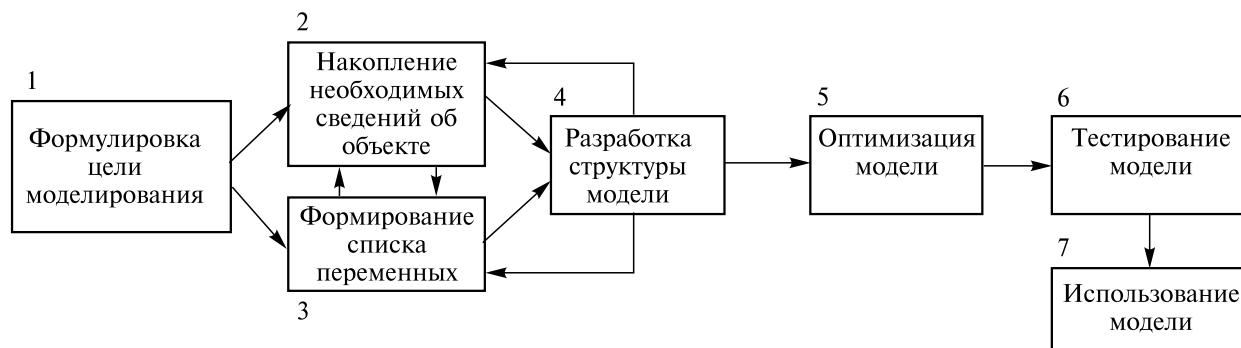


Рис. 1. Этапы моделирования.

соответствующей корректировки модели. В результате общие затраты на разработку оказываются заметно больше, чем в первом случае, а полученная модель становится не самым подходящим инструментом для достижения цели.

Исходя из опыта, можно сказать, что наилучший эффект достигается при выборе первого подхода как основного, но с использованием по мере необходимости элементов второго.

Составление списка контролируемых переменных. Сформулированные на первом этапе цели моделирования и собранные на втором информационные материалы позволяют перейти к третьему этапу, когда составляется список контролируемых переменных. Прежде всего в него включаются: характеристики, с использованием которых формализуется описание цели моделирования; переменные, обеспеченные данными измерений или другими оценками их значений; параметры, необходимые для формулировки теоретических или гипотетических положений, которые предполагается использовать при построении модели.

Для каждой включаемой в список переменной необходимо определить ее основные свойства: наименование, тип, единицу измерения, область изменчивости, точность контроля значений.

Наименование представляет собой краткое словесное описание, поясняющее содержательный смысл переменной. Дополнительно к словесному описанию наименование может включать ее краткое символическое обозначение.

Существует два основных типа переменных — количественные и качественные. Значение количественной переменной выражается числом, значение качественной — текстовым описанием, рисунком или каким-либо другим поясняющим его смысл способом.

Единицы измерения выбираются таким образом, чтобы удобно было записывать числовые значения количественных переменных. Определения единиц измерения для качественных переменных не требуется.

Для каждой используемой в модели характеристики нужно определить область ее изменчивости. Для качественной переменной это перечень всех ее значений. Для количественного параметра область изменчивости задается положением ее нижней и верхней границ.

При разработке информационно-вероятностной модели необходимо, чтобы число различаемых значений всех контролируемых характеристик объекта было конечным. Для качественных переменных это условие выполняется почти автоматически, поскольку составить бесконечный перечень описаний их различных значений практически невозможно. Однако, чем больше различаемых значений качественной переменной, тем выше точность ее контроля. Для количественных характеристик проблема ограничения числа различаемых значений решается подходящей дискретизацией областей изменчивости переменных (рис. 2, *a* и *b*) или путем задания конечного размаха этих областей и введением характерной точности (ошибки) контроля значений.

Если в качестве способа ограничения числа различаемых значений выбирается дискретизация количественных переменных, то их свойства становятся похожими на свойства качественных. Так, для случая *b* (см. рис. 2) вместо использования двух количественных характеристик можно использовать одну качественную.

Говоря о точности контроля задаваемых дискретно количественных переменных, можно связывать ее или с размером дискретных элементов, или с их общим количеством. При выборе второго способа ограничения множества различаемых значений задаваемая точность контроля количественных переменных определяется минимальной величиной модуля разности между двумя соседними значениями переменной, при которой они еще различаются. Например, если точность контроля переменной равняется единице, то два любых значения этой переменной, модуль разности которых меньше единицы, считаются одинаковыми. Понятно, что точность контроля в этом случае может быть как постоянной, так и меняющейся по области изменчивости рассматриваемой характеристики.

Перечень переменных, обладающих описанными выше свойствами, задает информационную емкость модели. Ее величина определяется количеством различных состояний объекта, которые могут быть отражены в модели. Чем больше информационная емкость, тем сложнее и информативнее может быть модель.

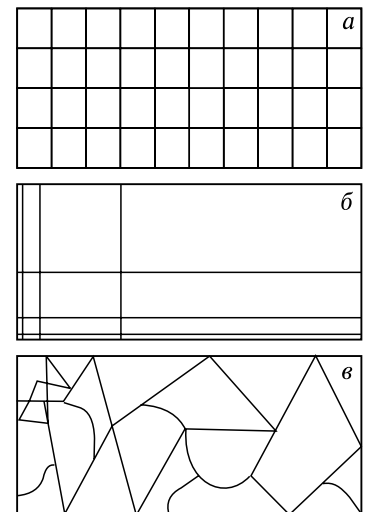


Рис. 2. Варианты дискретизации непрерывной прямоугольной области изменчивости двух переменных.

a — независимая по каждой из переменных, равномерная; *b* — независимая по каждой из переменных, неравномерная; *v* — совместная по обоим переменным, произвольная.

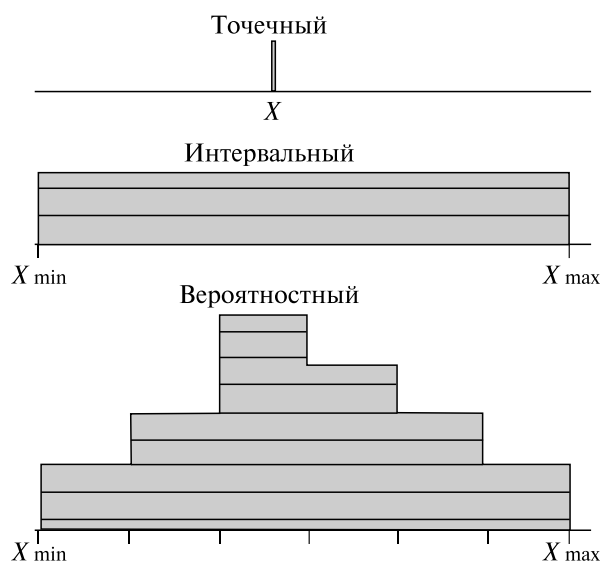


Рис. 3. Различные способы описания оценки значения переменной.

Выбранная дискретизация пространства переменных задает максимальную точность их контроля в модели. Информацию же о значениях аналогичных характеристик моделируемого объекта содержат их оценки, точность которых обычно меньше точности контроля переменных в модели. Для учета приближенности оценок применяются различные приемы. Среди них можно выделить три основных способа описания значений переменных: точечный, интервальный и вероятностный (рис. 3).

При наиболее распространенном точечном способе оценка описывается одним числом, несущим информацию о значении переменной. Точность такой оценки игнорируется или приравнивается

к некоторой стандартной величине. Например, в нашем случае она считается не меньше выбранной точности контроля переменной в модели.

При интервальном способе [5] используются два числа, определяющие границы интервала, внутри которого достоверно или с принятой по умолчанию вероятностью находится описываемое значение. Близким ему по смыслу является другой широко распространенный способ двухпараметрического описания, когда в явном виде приводятся точечная оценка и ее ошибка.

При вероятностном способе [6] значение переменной описывается с помощью функции плотности распределения вероятности. В теории нечетких множеств [7] функцию, предназначенную для той же цели, принято называть функцией принадлежности. Вместо фактической функции плотности вероятности на практике часто используется ее оценка, определяемая выбранным способом аппроксимации этой функции и набором ряда параметров. При информационно-вероятностном моделировании наряду с заданием плотности распределения, описывающей приближенное значение или характер изменчивости переменной, вводится показатель, определяющий меру доверия к такой оценке.

Точечные и вероятностные оценки применимы как к количественным, так и к качественным переменным, интервальные же используются для описания приближенных значений только количественных характеристик. Наибольшей общностью и информативностью обладают вероятностные оценки, а точечная и интервальная могут рассматриваться как частные случаи вероятностной.

Разработка структуры модели. На этом этапе формируется система ограничений на возможные значения контролируемых переменных, логически вытекающих из исходных данных. При построении информационно-вероятностной модели эта система строится в виде оценки совместного распределения вероятностей значений переменных модели, определенного в пространстве их изменчивости. Информативность (количество информации об объекте) такой модели определяется как разность между энтропией равномерного распределения и распределения, устанавливающего ограничения на возможные значения переменных. Способы преобразования исходных данных в оценку совместного распределения различны и представляют собой самостоятельные, зачастую достаточно сложные и нетривиальные задачи. Однако можно отметить некоторые общие черты процедуры построения оценки совместного распределения.

На ее начальном этапе из всех собранных данных выбираются те, которые могут рассматриваться как оценки отдельных значений контролируемых переменных или как оценки ограничений на множество их возможных значений. Способы описания оценок значений приводились ранее. Для описания ограничений на множество значений используются обычные математические соотношения равенства или неравенства между переменными или те или иные распределения вероятностей. Если среди исходных данных есть различные оценки одних и тех же значений переменных, то такие материалы следует объединить в одну более точную или достоверную оценку [6]. Подобная операция применяется и к различным оценкам одних и тех же ограничений на возможные значения переменных. В результате общее количество оценок сокращается, но далее каждая из них рассматривается как независимая.

Для того чтобы на основании данных о значениях переменных можно было сформировать оценку совместного распределения, их нужно упорядочить в матрицу совместных реализаций. Обычно в строки такой матрицы группируются значения переменных, относящихся к одному и тому же объекту

или к одному и тому же временному интервалу. При соответствующей статистической обработке построенной матрицы можно получить оценку искомого совместного распределения. Эта рутинная операция по преобразованию информации алгоритмируется в общем виде для произвольных исходных данных [6, 8, 9].

Вторую оценку совместного распределения можно получить путем объединения ограничений на возможные значения переменных. Для этого все соотношения между переменными разделяются на совместные и несовместные. Далее совместные ограничения объединяются путем логического умножения, а несовместные — путем логического сложения. Для реализации этих операций применяются, соответственно, правила умножения и сложения вероятностей [10], в результате чего исходная информация с учетом ее достоверности трансформируется в оценку совместного распределения переменных. Поскольку общего алгоритма для такой процедуры пока еще не создано, то для ее реализации требуются творческие усилия разработчика модели. Приближенным аналогом этой операции в математическом моделировании является составление системы уравнений или неравенств.

Две полученные оценки распределения вероятностей (на основе данных о наблюдавшихся значениях переменных и информации об ограничениях их возможных значений) могут быть использованы для синтеза результирующей оценки, но способ этого синтеза должен быть выбран с учетом свойств информации, содержащейся в каждой из оценок. Возможен подбор наиболее подходящего способа в соответствии с целью моделирования, но эту операцию лучше отнести к следующему этапу — оптимизации модели.

Оптимизация модели. Этот процесс ориентирован на получение наилучшей в некотором смысле оценки совместного распределения контролируемых переменных или специально выбранных его характеристик. Постановка оптимизационной задачи требует задания критерия оптимальности модели. Он может вытекать из цели работы или быть некоторым дополнительным требованием, предъявляемым к оптимизируемой оценке. При поиске точки экстремума заданного критерия могут варьироваться различные характеристики модели, например — значения постоянных коэффициентов в математических соотношениях между переменными, математическая структура этих соотношений, число и точность контроля переменных модели.

Наиболее известна и распространена первая из этих задач. Она обычно возникает при аппроксимации материалов измерений заданным математическим соотношением или при необходимости определения по экспериментальным данным неизвестных постоянных параметров в математической формулировке теоретической модели. При информационно-вероятностном моделировании такой задачей может быть минимизация расхождений между двумя оценками совместного распределения переменных, полученными на предыдущем этапе.

Вторая оптимизационная задача включает в себя первую, но кроме значений постоянных параметров в ней варьируется и математическая структура подбираемых соотношений между переменными. Например, выбирается лучшая из множества формул (математических гипотез), каждая из которых определена с точностью до одного и того же числа неизвестных параметров, наиболее приемлемо аппроксимирующая имеющиеся данные о совместных значениях характеристик.

Характерная задача третьего типа — нахождение оптимального числа квантилей гистограммы как оценки неизвестного распределения вероятностей по выборке фиксированного объема. Существование такого оптимума обусловлено тем, что конечная выборка данных содержит определенное количество эмпирической информации об искомом распределении. Подбор точности задания переменной (числа квантилей) позволяет согласовать информационную емкость модели с используемыми для ее построения данными. При малом числе квантилей значительная часть полезной информации теряется, при большом — ее оказывается недостаточно, и достоверность оценки распределения становится недопустимо низкой. В нашем случае количество информации об искомом распределении определяется содержанием, количеством и качеством (точностью и достоверностью) исходных данных, а число квантилей гистограммы — числом различимых значений переменных модели. Следовательно, варьируя тем или иным образом последнюю характеристику, можно подобрать оценку совместного распределения, которая наиболее полно соответствует цели моделирования.

Другой пример оптимизационной задачи третьего типа возникает в связи с проблемой поиска взаимосвязей между переменными, скрытых в данных их измерений [9]. При ее решении для каждой из переменных, рассматриваемых как зависимые, определяется наилучший список аргументов, по значениям которых наиболее точно предсказывается значение функции. Возможны и другие варианты задач оптимизации моделей.

Рассмотрим пример решения оптимизационной задачи (в рамках построения прогностической модели стока в Байкал [6]), в которой подбирались наилучшая по предсказательной способности регрессионная зависимость. Варьируемыми характеристиками в этой задаче были: число предикто-

**Определение критерия предсказательной способности
регрессионных моделей в зависимости от их сложности**

Точность задания предикторов, % от размаха области их изменчивости	Количество используемых в модели предикторов					
	2	3	4	5	6	7
100	0,094	0,106	0,131	0,139	0,146	0,151
50	0,119	0,206	0,231	0,214	0,194	0,183
30	0,071	0,134	0,174	0,108	0,098	0,085
20	0,040	0,066	0,056	0,048	0,043	0,039
10	0,012	0,014	0,011	0,010	0,010	0,010
1	0	0	0	0	0	0

зируемой переменной. Как показывают результаты расчетов (см. таблицу), в рассмотренном примере действительно наблюдается оптимум сложности модели, когда достигается максимальная точность прогноза. В данном случае — это модель с четырьмя предикторами, каждый из которых имеет разрешение в две-три градации.

Тестирование модели. При создании любой модели не исключены непреднамеренные ошибки, совершаемые в процессе ее построения. Чтобы уменьшить вероятность их появления, целесообразно произвести тестирование модели. Можно выделить три основных способа выполнения этой операции — проверку логики построения модели, восстановление исходных данных, проверку предсказательной способности модели на материалах, не использовавшихся для ее построения.

Для реализации первого способа лучше всего детально изложить процесс построения модели для критически настроенных, но объективных экспертов. Применимость второго способа ограничивается тем, что при прямом преобразовании данных (построение модели) и обратном (восстановление исходных материалов) часть информации может теряться в силу свойств этих преобразований. Поэтому, за исключением отдельных ситуаций, речь может идти только о приблизительном совпадении исходных и восстановленных данных.

Для реализации третьего способа исходные данные разделяются на обучающую и контрольную выборки. Модель строится на основе обучающих данных, а контрольные используются для ее проверки. При такой технологии модели получаются более простыми, так как основываются на меньшем количестве информации, но часть из них (даже если они выдерживают первый и второй способ проверки) оказывается неадекватной по третьему тесту. Степень же доверия к моделям, выдержавшим все три испытания, повышается.

Использование модели. Уже упоминалось, что применение моделей в большинстве случаев связано с оценкой условных значений или распределений переменных. При использовании математических моделей такое действие требует применения множества различных методов (аналитических и численных), во многом специфических для каждой конкретной задачи. Удобство использования информационно-вероятностной модели, задаваемой оценкой совместного распределения, состоит в том, что условное распределение для любой контролируемой переменной находится по стандартной для всех случаев формуле Байеса [10, 11].

Причем при расчете на компьютере требуемой условной характеристики не нужно вычислять все значения совместного распределения вероятностей. Достаточно путем преобразования исходных данных найти только те его значения, которые непосредственно используются в конкретном случае. Это существенно сокращает потребности в вычислительных ресурсах и позволяет эффективно находить конструктивные решения различных географических задач [6, 9].

СПИСОК ЛИТЕРАТУРЫ

1. Дружинин И. П., Хамьянова Н. В., Лобановская Ю. А. Прогноз гидрометеорологических элементов. — Новосибирск: Наука, 1977.
2. Абасов Н. В., Бережных Т. В., Резников А. П. Долгосрочное прогнозирование природообусловленных факторов в энергетике // Системные исследования проблем энергетике. — Новосибирск: Наука, 2000.
3. Виноградов Ю. Б. Математическое моделирование процессов формирования стока. — Л.: Гидрометеиздат, 1988.
4. Орлов Ю. К. Основы теории доказательств в уголовном процессе. — М.: Проспект, 2000.
5. Алефельд Г., Херцбергер Ю. Введение в интервальные вычисления. — М.: Мир, 1987.

6. **Игнатов А. В.** Модели и оптимизационные задачи в проблемах природопользования в Байкальском регионе. — Новосибирск: Изд-во СО РАН, 2000.
7. **Аверкин А. Н.** Нечеткие множества в моделях управления и искусственного интеллекта / Под ред. Д. А. Поспелова. — М.: Наука, 1986.
8. **Игнатов А. В.** Оценка полной экспериментальной информации о связи показателей состояния природных объектов // Модели и методы оценки антропогенных изменений геосистем. — Новосибирск: Наука, 1986.
9. **Игнатов А. В.** Опыт вероятностного моделирования и анализа взаимозависимости многомерных географических данных // География и природ. ресурсы. — 1996. — № 4.
10. **Вентцель Е. С.** Теория вероятностей. — М.: Гос. изд-во физ.-мат. лит-ры, 1962.
11. **Корн Г., Корн Т.** Справочник по математике. — М.: Наука, 1977.

*Институт географии СО РАН,
Иркутск*

*Поступила в редакцию
5 сентября 2005 г.*