

МОНИТОРИНГ ПРИРОДНЫХ СИСТЕМ

УДК 519.25:551.326:551.14

МОНИТОРИНГ СОСТОЯНИЙ МНОГОКОМПОНЕНТНЫХ ОБЪЕКТОВ В ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ

А.Н. Четырбоцкий

Дальневосточный геологический институт ДВО РАН, г. Владивосток

Рассмотрены методические подходы для решения задач выявления состояний объектов, которые заданы выборками многомерных наблюдений. Применение подходов иллюстрируется решением типичных задач. Полученные результаты и предлагаемые методики могут быть использованы для исследования реальных объектов в геологии, океанографии и других прикладных областях.

Ключевые слова: объект, наблюдения, параметрический портрет, выборка данных, многомерное шкалирование.

ВВЕДЕНИЕ

Для создания геоинформационных систем (ГИС) природных объектов следует провести определенный комплекс исследований. Ключевой этап таких исследований состоит в конструировании параметрического портрета или геоинформационной модели рассматриваемого объекта. Для этого надлежит произвести отбор ограниченного числа исходных показателей или их функций, на основании которых однозначным образом определяется состояние объекта. Понятно, что число таких состояний и топология распределения совокупности объектов в признаковом пространстве представляют собой важные сведения об их генезисе. Если объекты представлены многомерной выборкой распределений соответствующих показателей, то эффективное решение таких задач может быть найдено методами многомерного шкалирования [1, 3, 4]. Сущность этих методов состоит в следующем [6, 7]. Полагается, что отдельные объекты выборки соответствуют точкам исходного многомерного пространства (термины “наблюдение”, “объект” и точка многомерного пространства являются эквивалентными). На основании принятой в исследовании меры близости между парой точек рассчитывается матрица близости между отдельными объектами. Далее в пространстве существенно меньшей размерности выполняется поиск такого расположения образов этих точек, матрица бли-

зости между которыми как можно меньше отличается от матрицы близости между точками в исходном пространстве. В настоящем рассмотрении используется евклидова метрика. Если наблюдения выборки отражают пространственно-временные изменения объектов, то подобным образом выполняется мониторинг их состояний.

Актуальность решения указанных задач и формулировка предметной интерпретации полученных результатов обусловлена тем, что построение геоинформационной модели обычно выполняется на уровне эвристических представлений. Поэтому в некоторых реальных ситуациях имеют место случаи как отсутствия необходимых сведений, так и их избыток.

В работе предлагаются примеры решения практических задач, которые достаточно часто имеют место при построении ГИС [9, 10]. Так, задача построения агрегированных показателей на основании выборочных распределений позволяет выявить структуру и, стало быть, латентный механизм изучаемого явления. Формулирование следующей задачи состоит в выявлении временных состояний комплексного объекта, в качестве которого выступает ледяной покров Японского моря. В настоящем случае объект представлен выборкой многолетних десятилетних состояний в отдельных, достаточно однородных по своим физико-химическим свойствам районах акватории моря. Целесообразность рассмотрения здесь этой задачи обусловлена общностью ее постановки и показательностью решения.

ОЦЕНКА ВРЕМЕННЫХ СОСТОЯНИЙ ЛЕДЯНОГО ПОКРОВА ЯПОНСКОГО МОРЯ

Количественное представление морского ледяного покрова (ЛП) как объекта изучения обычно выполняется на основании усредненных за отдельные декады выборочных распределений сплоченности (отношение площади льда в районе к площади района), возрастного состава льдов (выраженные в баллах значения толщин покрова) и преобладающего размера отдельных льдин. Совокупность этих распределений характеризует набор состояний рассматриваемой динамической системы, где термин «система» непосредственно соответствует его классическому определению «...целое, составленное из частей; совокупность элементов, находящихся в отношениях и связях друг с другом, которая образует определенную целостность, единство» [5]. Иными словами – это иерархически упорядоченная и целенаправленно функционирующая совокупность большого числа взаимодействующих элементов. В качестве последних выступает ЛП отдельных районов акватории моря, а набор состояний системы или объекта настоящего рассмотрения определяется выборкой таких распределений. Здесь объект представлен выборкой многолетних декадных распределений площадей и толщин льда за период 1961–1989 гг. Каждое распределение характеризует ЛП отдельного района моря, где выделение районов было выполнено исходя из однородности, в определенном смысле, его участков. Материал для исследований предоставлен проф. В.В. Плотниковым (Тихоокеанский океанологический институт ДВО РАН).

Максимальная продолжительность эволюционного цикла объекта составляет 22 декады (начальной является 31-я декада года, а конечной – 16-я декада) [8]. Поскольку в исходных выборках значения характеристик ЛП приведены в баллах соответствующих шкал измерений, то для выполнения исследований была выполнена их оцифровка [11]. Принимаются обозначения: S_{dry} – сплоченность, H_{dry} – толщина и F_{dry} – преобладающий размер отдельных льдин покрова в d -й декаде ($d = 1 \div 22$) r -го района ($r = 1 \div 166$) в y -м году ($y = 1961 \div 1989$). Полагается также, что число районов ежегодного присутствия льда равно $N_r = 166$ (114 выделенных в [2] районов плюс 52 района, где присутствует припай). На рис. 1 представлена карта расположения и нумерация районов.

Состояние объекта определяется совокупностью выборок распределений $\{S_{dry}\}$, $\{H_{dry}\}$ и $\{F_{dry}\}$ для каждого района акватории моря. Вследствие высокой функциональной зависимости между S и F (коэффициент корреляции между наборами $\{S_{dry}\}$ и $\ln\{F_{dry}\}$ равен 0.946), состояние покрова определяется распределением $\{S_{dry}\}$ и распределением $\{H_{dry}\}$. Понятно, что из этих распределений можно сформировать распределение $\{(S \cdot H)_{dry}\}$, которое имеет физический смысл распределений объемов льда. Для удобства диапазон изменения этих чисел приведен к отрезку $[0,1]$. Оценка декадных состояний объекта выполняется на основании выборки 22 наблюдений (22 декады), каждое из которых насчитывает $N_d \cdot N_r = 4814$ элементов ($N_d = 29$ лет наблюдений и $N_r = 166$ районов). На рис. 2 приведен результат иерархической классификации и решение соответствующей задачи многомерного шкалиро-

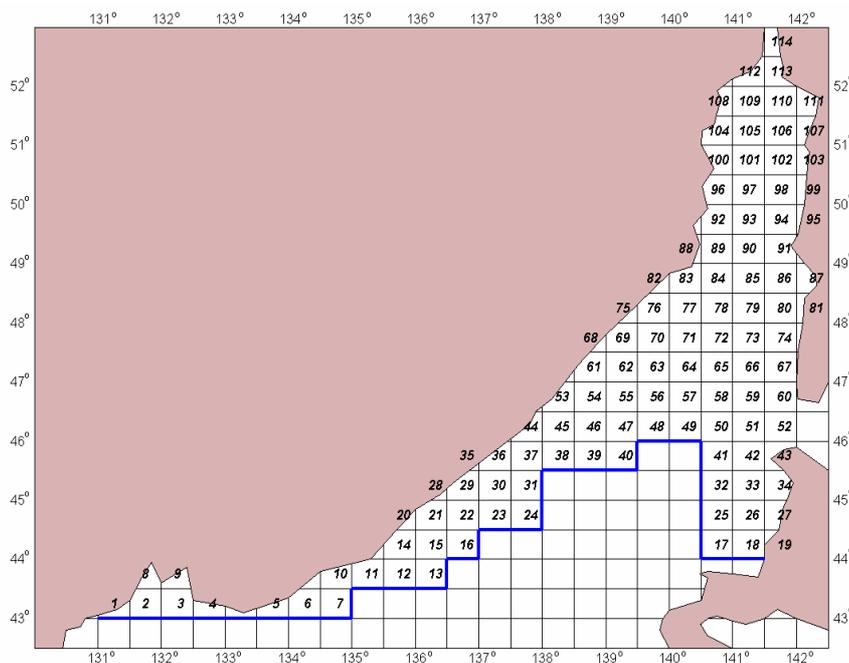


Рис. 1. Карта расположения и нумерация районов Японского моря [2], где отмечается присутствие льда (толстая линия указывает границу максимального распространения льда).

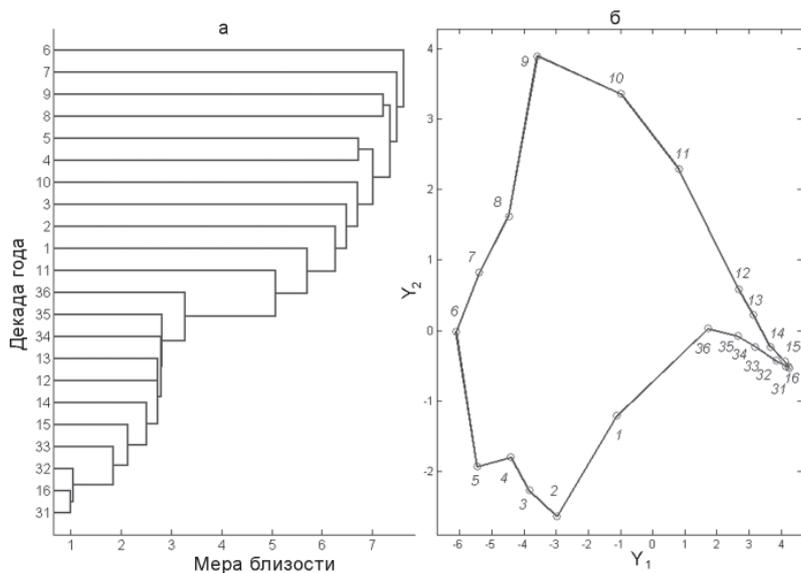


Рис. 2. Оценка декадных состояний ЛП Японского моря (выборку составляют наборы $\{(S \cdot H)_{dry}\}$).

а – результаты иерархической классификации; б – решение задачи многомерного шкалирования, где цифры указывают соответствующую декаду года.

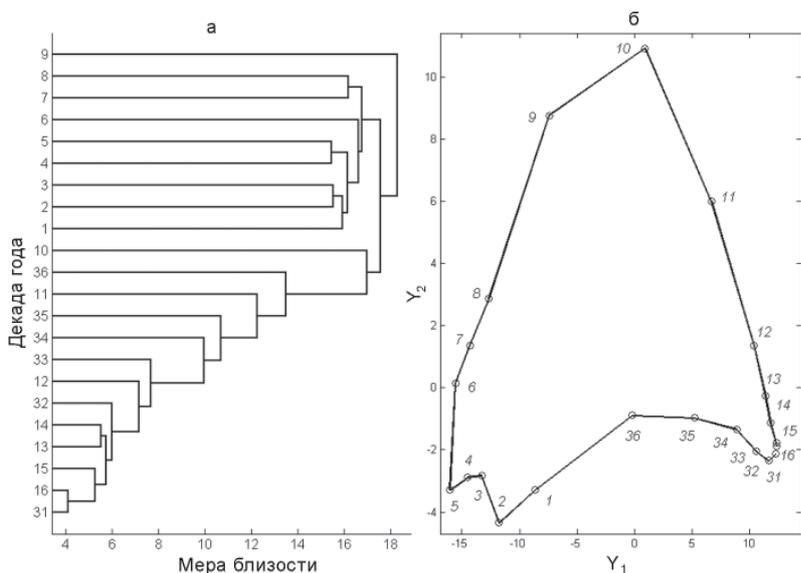


Рис. 3. Оценка декадных состояний ЛП Японского моря (выборку составляют наблюдения $\{S_{dry}, H_{dry}\}$).

а – результаты иерархической классификации; б – решение задачи многомерного шкалирования, где цифры указывают соответствующую декаду года.

вания (пространством образов исходных точек является плоскость $Y_1 Y_2$). В обоих случаях в качестве меры близости выступает евклидова метрика.

Конфигураций ветвей бинарного дерева на рис. 2а указывает на неразличимость распределений осенних и весенних этапов. Действительно, согласно результатам классификации, наблюдения этих декад (с 31 по 36 и с 11 по 16 декады) являются ветвями одного поддерева. Наблюдения зимних декад (с 1 по 10 декаду) образуют отдельные ветви. Окончания осеннего (36 декада) и зимнего (11 декада) этапов являются пограничными. Иерархические бинарные деревья характеризуют основную тенденцию распределения выборочных точек в пространстве $N_d \cdot N_r$ размерности. В этом смысле нумерация конечных ветвей дерева не позволяет оценить реальное пространственное распо-

ложение наблюдений, которые принадлежат его другим бинарным ветвям. Характер расположения образов точек на рис. 2б в большей степени отражает распределение наблюдений. Однако и в этом случае наблюдения осеннего и весеннего этапов не отличаются между собой. Необходимо отметить, что так называемый кофенетический коэффициент корреляции [1] (коэффициент корреляции между матрицами близости для случаев а) и б)) равен 0.930. Согласно этим результатам, в терминах объемов льда эволюция ЛП представляет собой обратимый термодинамический процесс: конечный весенний этап таяния покрова является обратным начальному осеннему этапу его формирования [11]. В этой ситуации состояния объекта для осенних и весенних декад не различаются между собой. Представленные на рис. 3 результаты аналогич-

ных построений для составной выборки признаков $\{S_{dry}, H_{dry}\}$ все же показывают их отличия.

В этом случае вектор наблюдений насчитывает $N_d \cdot N_r \cdot 2 = 9628$ элементов, а кофенетический коэффициент корреляции равен 0.758. В терминах площадей и толщин декадные состояния объекта отличаются между собой. Указанное положение допускает простую интерпретацию. Отличие состояний обусловлено следующим: при одном и том же объеме у льдин осеннего этапа формирования большая площадь и малая толщина, а у льдин весеннего этапа таяния малая площадь (вследствие быстрого таяния тонкого льда) и повышенная толщина. Таким образом, последующее математическое моделирование эволюции покрова следует выполнять в терминах площадей и толщин покрова.

ТИПИЗАЦИЯ ЛЕДЯНОГО ПОКРОВА РАЙОНОВ АКВАТОРИИ ЯПОНСКОГО МОРЯ

Начальный этап построения модели (логической, физической, математической и т.д.) системы произвольной природы состоит в упорядочивании и систематизации реальных фактов. Если система представлена выборкой наблюдений, то для выявления фактов и поиска латентных (скрытых, непосредственно не наблюдаемых) показателей, в пространстве которых они проявляются, в большинстве случаев следует выполнить типизацию (классификацию) исходных наблюдений. В настоящем случае число и состав кластеров является известным [2]: как правило, ледяной покров вследствие ледовых условий в районах акватории моря территориально разделяют на три части: Татарский пролив, район вдоль побережья Приморья от мыса Поворотного до мыса Белкина и залив Петра Великого). Вместе с тем, открытым является вопрос о действительном числе таких кластеров, а также характере проявления территориального разделения в пространстве метрических параметров выборочных состояний покрова (далее при упоминании понятия “состояние” следует иметь в виду его выборочность).

При решении задач необходимо учитывать территориальную неоднородность покрова льда, несовпадение дат первичного формирования ЛП в отдельных районах и продолжительностей периодов их эволюции. Данная ситуация обуславливает тот факт, что понятие “ледяного покрова усредненного района” является в большей мере фиктивным понятием исследуемой системы (реальными характеристиками системы являются средние за многолетний период значения параметров в отдельных декадах каждого района). Поэтому для решения задач целесообразным представляется использование методов, где от-

сутствует статистическая парадигма усреднения характеристик выборки. К числу таких методов относятся методы многомерного шкалирования. Решение задач многомерного шкалирования позволяет перейти от формы исходных данных типа «объект-объект» к форме типа “объект-свойство” [1]. В этом случае на основании объектов выборки автоматически формируется выборка их образов существенно меньшего числа агрегированных показателей, в пространстве которых легко выполнить дифференциацию объектов исходной выборки.

Оценивание адекватности (валидация) конфигурации образов объектов в пространстве агрегированных показателей конфигурации объектов в исходном пространстве выполняется на основании так называемого кофенетического коэффициента корреляции

$$c = \frac{\sum_{i < j} (R_{ij} - \bar{R}) \cdot (Z_{ij} - \bar{Z})}{[\sum_{i < j} (R_{ij} - \bar{R})^2 \sum_{i < j} (Z_{ij} - \bar{Z})^2]}^{1/2}, \quad (1)$$

где R_{ij}, Z_{ij} – значения мер близости между i и j объектами в исходном пространстве и пространстве их образов, где $i = 1 \div K-1$ и $j = i + 1 \div K$; K – общее число объектов; \bar{R}, \bar{Z} – соответствующие средние. Понятно, что близость c к единице указывает на соответствие между указанными конфигурациями. В приведенном ниже исследовании посредством c_1, c_2 и c_3 обозначены рассчитанные по (1) коэффициенты для первого агрегированного показателя, первых двух агрегированных показателей и первых трех агрегированных показателей. Верхними индексами будем обозначать соответствующую признаковую систему. Для удобства интерпретации результатов изменение F приведено к отрезку $[0, 1]$.

В данном случае $K = 166$ объектов: 114 районов моря, из которых в 52 прибрежных районах отмечаются льды открытого моря и отдельно льды припая. Признаками объектов являются среднедекадные значения сплоченности (22 признака), толщины покрова (22 признака) и преобладающего размера льдин (22 признака) в отдельных 29 годах наблюдений. Поскольку анализу подлежат проявления особенностей территориального расположения отдельных районов, то признаковая система рассматриваемых объектов может насчитывать от $22 \cdot 29 = 638$ до $3 \cdot 22 \cdot 29 = 1914$ элементов. На рис. 4 представлены решения задач многомерного шкалирования, где маркировка осей указывает на соответствующий признаковый набор (верхние индексы) и порядковый номер агрегированного показателя (нижний индекс); большими цифрами указаны положения центров предварительно выделенных нами групп точек. Выделение групп было вы-

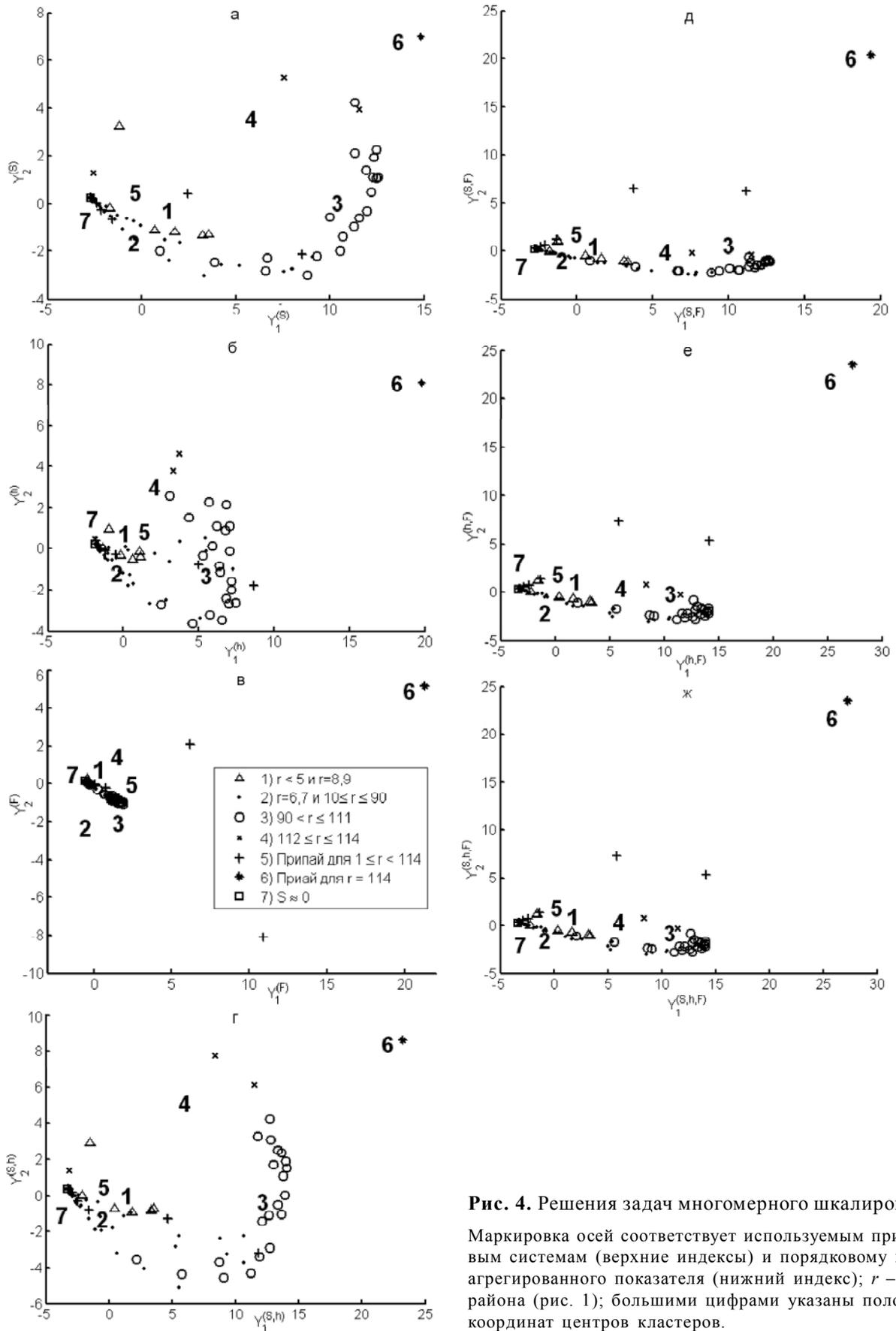


Рис. 4. Решения задач многомерного шкалирования.

Маркировка осей соответствует используемым признаковым системам (верхние индексы) и порядковому номеру агрегированного показателя (нижний индекс); r – номер района (рис. 1); большими цифрами указаны положения координат центров кластеров.

полнено таким образом, что ледяные покровы отдельных районов со схожими ледовыми режимами составляли одну группу образований льда (каждый район акватории определяется своим номером на рис. 1). Согласно этому требованию, было выделено 7 групп районов: ЛП открытой части районов с 1 по 4, 8 и 9, которые составляют залив Петра Великого (1); ЛП открытой части районов с номерами 6, 7 и $9 < r \leq 90$ (2); ЛП открытой части районов $90 < r < 112$ (3); ЛП открытой части районов Татарского пролива $112 \leq r \leq 114$ (4); припай районов с 1 по 113 (5); припай 114-го района (6); ЛП тех районов, где за период наблюдений отмечаются незначительные и нерегулярные появления льдов.

Если учитывается сплоченность S ледяных покровов выделенных участков моря (размерность пространства $22 \cdot 29 = 638$), то конфигурация образов наблюдений на плоскости первых двух агрегированных показателей (рис. 4а) и анализ отмеченных выше корреляций показывает следующее. Проявляется явное отличие припая в районе пролива Невельского (114 район) от образований льда в других районах акватории моря, что обусловлено узостью пролива (расстояние между берегами чуть более 7 км). Остальные выборочные точки распределены в узкой области данного признакового пространства, где отсутствуют изолированные сгущения точек. Размерность этой области существенно меньше размерности признакового пространства. Поэтому для представления этого распределения достаточно ограничиться только небольшим числом агрегированных показателей. В данном случае достаточно ограничиться двумя показателями. Действительно, выборочная оценка кофенетического коэффициента корреляции $c_1^{(S)}$ между матрицами близости наблюдений в исходном пространстве и построенной только на основании первого агрегированного показателя $Y_1^{(S)}$ равна $c_1^{(S)} = 0.961$, а $c_2^{(S)} = 0.975$ и $c_3^{(S)} = 0.981$. Согласно распределению образов выборочных точек, имеет место плавный переход от ЛП зал. Петра Великого и припая в большей части районов акватории ко льдам районов Татарского пролива и припая пролива Невельского. Отсутствие изолированных сгущений в конфигурации указывает на тот факт, что состояния ледяного покрова в отдельных районах на акватории моря следует рассматривать с позиций существования единого механизма их генерации.

Распределение выборочных точек на рис. 4б (учитываются толщины льда h , размерность пространства $22 \cdot 29 = 638$) в большей степени соответствуют относительно равномерному заполнению шарообразной области данного пространства. Исключение составляют толщина припая 114-го района, глубина которого

не превышает 10 м. Следовательно, распределение толщин не зависит от принадлежности образований льда ко льдам открытого моря или льдам припая. Для представления этого заполнения можно ограничиться двумя агрегированными показателями: $c_1^{(h)} = 0.925$, $c_2^{(h)} = 0.946$ и $c_3^{(h)} = 0.957$. Также как и в предыдущем случае, данная ситуация указывает на единый механизм генерации состояний ледяного покрова в отдельных районах моря.

Конфигурация образов выборочных точек на рис. 4в (учитываются преобладающие размеры льдин F ЛП участков моря, размерность пространства $22 \cdot 29 = 638$) так же, как и в случае на рис. 4а, находится в узкой области данного признакового пространства. Факт явного отличия F для группы с номером 6 обусловлен узостью 114-го района. Здесь также отсутствуют изолированные сгущения точек. Размерность этой области существенно меньше размерности признакового пространства. Для представления конфигурации выборочных точек можно ограничиться двумя агрегированными показателями: $c_1^{(F)} = 0.970$, $c_2^{(F)} = 0.973$ и $c_3^{(F)} = 0.987$.

Конфигурация образов выборочных точек на рис. 4г (учитываются сплоченности S и толщины покрова h , размерность пространства $2 \cdot 22 \cdot 29 = 1276$) соответствует конфигурации точек на рис. 4а. Поэтому в пространстве сплоченности и толщин состояния ледяного покрова в отдельных районах акватории моря следует рассматривать с позиций единого механизма их генерации. Для представления конфигурации выборочных точек можно ограничиться двумя агрегированными показателями: $c_1^{(S,h)} = 0.944$, $c_2^{(S,h)} = 0.962$ и $c_3^{(S,h)} = 0.969$.

Конфигурация образов выборочных точек на рис. 4д (учитываются сплоченности S и преобладающие размеры льдин F ЛП участков моря, размерность пространства $2 \cdot 22 \cdot 29 = 1276$) подобна конфигурации на рис. 4в. Для представления конфигурации выборочных точек можно ограничиться двумя агрегированными показателями: $c_1^{(S,F)} = 0.945$, $c_2^{(S,F)} = 0.967$ и $c_3^{(S,F)} = 0.979$.

Конфигурация образов выборочных точек на рис. 4е (учитываются толщины покрова h и преобладающие толщины льдин F ЛП участков моря, размерность пространства $2 \cdot 22 \cdot 29 = 1276$) соответствует конфигурации точек на рис. 4б. Поэтому в пространстве толщин покрова и преобладающего размера льдин распределение толщин не зависит от принадлежности образований льда ко льдам открытого моря или льдам припая. Для представления конфигурации выборочных точек можно ограничиться двумя агрегированными показателями: $c_1^{(h,F)} = 0.936$, $c_2^{(h,F)} = 0.945$ и $c_3^{(h,F)} = 0.963$.

Конфигурация образов выборочных точек на рис. 4ж (учитываются сплоченности S , толщины покрова h и преобладающие толщины льдин F ЛПІ участков моря, размерность пространства $3 \cdot 22 \cdot 29 = 1914$) соответствует конфигурации точек на рис. 4а. Поэтому в пространстве сплоченности, толщины покрова и преобладающие размеры льдин состояния ледяного покрова моря в отдельных районах акватории следует рассматривать с позиций единого механизма их генерации. Для представления конфигурации выборочных точек можно ограничиться двумя агрегированными показателями: $c_1^{(S,h,F)} = 0.940$, $c_2^{(S,h,F)} = 0.953$ и $c_3^{(S,h,F)} = 0.970$.

На основании выявленных фактов можно сделать следующий вывод: в каждом рассмотренном признаковом пространстве конфигурация выборочных точек находится в узкой области, где отсутствуют изолированные сгущения точек. Размерность этой области существенно меньше размерности признакового пространства. Исключение составляет образование льда припая пролива Невельского, что обусловлено его узостью и мелководностью. В различных конфигурациях отмечается устойчивое сохранение порядка следования выборочных точек. Данные обстоятельства указывают на единый механизм генерации состояний ледяного покрова в отдельных районах моря.

ЗАКЛЮЧЕНИЕ

Для иллюстрации рассмотренного здесь подхода приведены два примера. Для обоих случаев логическая цепочка рассуждений одинакова: на основании статистически представительной выборки следует выполнить оценку числа сгущений и конфигурации точек (наблюдений) в многомерном пространстве признаков. Назначение отдельных вычислительных процедур звеньев цепочки состоит в тестировании статистической значимости и предметной интер-

претации полученных результатов. Чтобы полученные результаты имели статистическую значимость, исходная выборка должна обладать значительным объемом и отражать основные особенности рассматриваемого явления.

ЛИТЕРАТУРА

1. Дэйвисон М. Многомерное шкалирование: Методы наглядного представления данных, М.: Финансы и статистика, 1988. 254 с.
2. Плотников В.В. Изменчивость ледовых условий Дальневосточных морей России и их прогноз. Владивосток: Дальнаука, 2002, 172 с.
3. Стронгин Р.Г. Численные методы в многоэкстремальных задачах (информационно-статистические алгоритмы), М.: Наука, 1978, 240 с.
4. Терехина А.Ю. Методы многомерного шкалирования визуализации данных(обзор) // Автоматика и телемеханика, 1973. № 7. С. 80–94.
5. Философский энциклопедический словарь / Гл. ред. Ильичев Л.Ф., Федосеев П.Н., Ковалев С.М., Панов В.Г. М.: Сов. Энциклопедия, 1983. 840 с.
6. Четырбоцкий А.Н. Методы и алгоритмы решения задач снижения размерности пространства описания. Владивосток, 1991, 95 с.
7. Четырбоцкий А.Н. Пакет прикладных программ для решения задач снижения размерности пространства описания // Управляющие системы и машины, Киев: Наук. думка, 1991. Т. 1(111). С. 103–106.
8. Четырбоцкий А.Н. Формирование и разрушение ледяного покрова замерзающих морей // Криосфера Земли. 2005. Т. 9, № 3. С. 45–53.
9. Четырбоцкий А.Н. Прогноз ледового режима на акватории Японского моря (математическая модель) // Информационные технологии. 2005. № 8. С. 9–14.
10. Четырбоцкий А.Н. Технология построения геоинформационных систем для ледяного покрова замерзающих морей (на примере ГИС “Ледяной покров Японского моря”) // Информационные технологии. 2005. № 11. С. 17–23.
11. Четырбоцкий А.Н. Статистический анализ параметров состояний ледяного покрова Японского моря и математическое моделирование его эволюции // Водные ресурсы. 2006. Т. 33, № 3. С. 289–299.

Поступила в редакцию 19 февраля 2006 г.

Рекомендована к печати А.И. Ханчуком

A.N. Chetyrbotsky

Monitoring of conditions of multicomponent objects in geoinformation systems

Approaches for the solution of problems of revealing conditions of objects are considered which are specified by samples of multivariate observations. Application of the approaches is illustrated by the solution of typical problems. The results obtained and the suggested techniques can be used for investigation of real objects in geology, oceanography, and in other applied fields.

Key words: object, observations, parametrical portrait, data sampling, multivariate scaling.