

## **Глава IV. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ**

### **ВВЕДЕНИЕ**

В первом параграфе приводятся результаты использования метода при факторном и кластерном анализе административных территорий Красноярского края по показателям здоровья и здравоохранения.

Результаты тестирования метода по таблице выборов президентов США приводятся во втором параграфе.

Третий параграф посвящен использованию программных продуктов FАMaster и ModelAnalyzer в задачах гелиофизики.

Использование метода моделирования данных с пробелами многообразиями малой размерности при обнаружении факторов, влияющих на течение и прогноз заболевания у больных со сложными нарушениями ритма и проводимости сердца, описано в четвертом параграфе.

### **IV.1. ФАКТОРНЫЙ И КЛАСТЕРНЫЙ АНАЛИЗ**

#### **АДМИНИСТРАТИВНЫХ ТЕРРИТОРИЙ КРАСНОЯРСКОГО КРАЯ ПО ПОКАЗАТЕЛЯМ ЗДОРОВЬЯ И ЗДРАВООХРАНЕНИЯ**

Эффективное централизованное управление здравоохранением такого крупного региона, как Красноярский край, представляет собой трудную задачу, в решении которой необходимо учитывать множество различающихся факторов, используя дифференцированный подход к планированию мероприятий и затрат.

Для эффективного управления здравоохранением актуальной является задача типизации отдельных регионов края с учетом взаимосвязей между различными факторами, с большей или меньшей силой оказывающими влияние на планирование и результат управленческих решений. Имеющиеся статистические данные, безусловно, помогают в принятии управленческих решений, однако обилие показателей далеко не всегда позволяет осмыслить ситуацию в целом.

Целью данного исследования стала типизация и выделение относительно похожих групп в системе из 49 регионов (административных территорий) Красноярского края с помощью кластерного анализа.

В расчетах использовались восходящая и нисходящая иерархические классификации по методу полной связи (дальнего соседа), а так же разбиение на классы с использованием метода динамических ядер [32]. Для анализа полученных разбиений использовались методы итерационного моделирования данных линейными и самоорганизующимися многообразиями малой размерности [14]. Все расчеты проводились с помощью программы "Model Analyzer 2.0", разработанной в рамках данной работы.

В системе здравоохранения края устанавливается деление всей территории края на 49 административных единиц (7 городов и 42 района). Для их кластеризации были использованы относительные показатели за 1999 год,

группирующиеся в 2 основных блока: 1) "Здоровье населения" (общая и первичная заболеваемость по возрастным группам и по классам заболеваний – 849 показателей); 2) "Кадры" (обеспеченность врачами разных специальностей) – 100 показателей). Кроме того, использовался параметр "Затраты на одного жителя в системе здравоохранения". Таким образом, общее число показателей составило 950.

Для анализа данных первого и второго блоков были использованы методы итерационного моделирования данных линейными и самоорганизующимися многообразиями малой размерности [14]. Цель применения данных методов заключалась в упорядочивании объектов (регионов) по главной компоненте в пространстве признаков каждого блока. Это позволило отсортировать регионы на условной шкале "хорошо – плохо" по каждому блоку данных. Согласно расположению регионов в пространстве признаков вдоль главной компоненты регионы были упорядочены отдельно по 1 и 2 блокам. Таким образом, для каждого региона были получены значения на 3 шкалах: 1) уровень здоровья населения; 2) обеспеченность кадрами; 3) затраты на здравоохранение в расчете на одного жителя (эти данные имелись исходно).

Наибольший интерес представляет соотношение затрат на одного жителя и уровнем здоровья населения (рис. 4.1), что, по нашему мнению, достаточно четко может характеризовать эффективность управления здравоохранением региона.

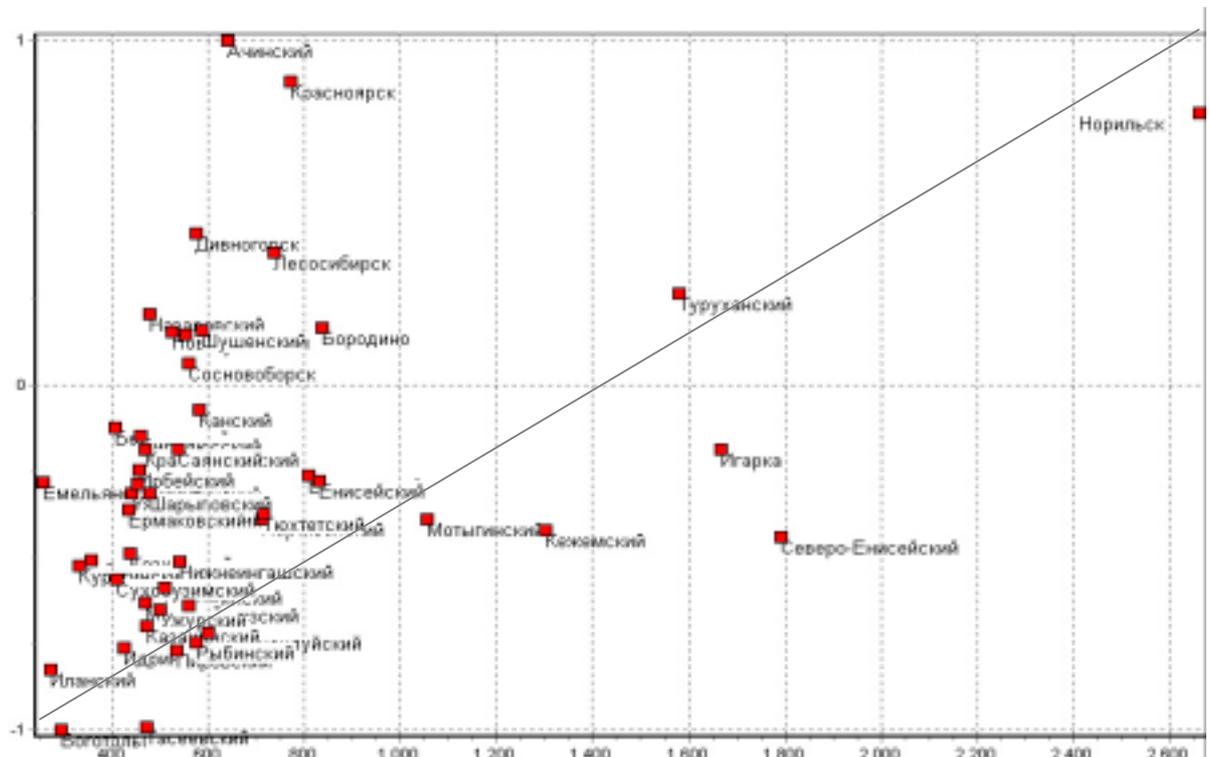


Рис. 4.1. Взаимосвязь между затратами на 1 жителя (ось X) и здоровьем населения (ось Y).

По всем трем показателям была проведена кластеризация регионов. Оптимальным стало разделение регионов на 5 кластеров (таб. 4.1 и рис. 4.2).

Таблица 4.1. Кластеризация регионов по трем показателям (совокупный показатель "Здоровье", совокупный показатель "Кадры", показатель "Затраты на 1 жителя").

№	Кол-во объектов	Объекты
1	10	г. Бородино, г. Дивногорск, г. Лесосибирск, г. Сосновоборск, Енисейский, Канский, Минусинский, Назаровский, Новоселовский, Шушенский р-ны
2	5	г. Игарка, Кежемский, Мотыгинский, Северо-Енисейский, Туруханский р-ны.
3	2	г. Красноярск, Ачинский р-н.
4	1	г. Норильск.
5	31	Абанский, Балахтинский, Березовский, Бирилюсский, Боготольский, Богучанский, Большемуртинский, Большеулуйский, Дзержинский, Емельяновский, Ермаковский, Идринский, Иланский, Ирбейский, Казачинский, Каратузский, Козульский, Краснотуранский, Курагинский, Манский, Нижнеингашский, Партизанский, Пировский, Рыбинский, Саянский, Сухобузимский, Тасеевский, Тюхтетский, Ужурский, Уярский, Шарыповский р-ны

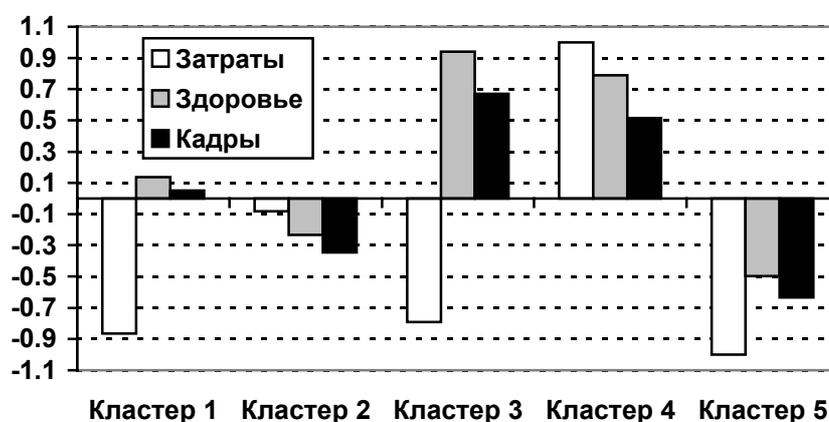


Рис. 4.2. Значения кластерных средних по совокупным показателям ("Здоровье" и "Кадры") и по показателю затрат на 1 жителя. Для удобства восприятия диаграммы значения показателя «Затраты на 1 жителя» нормированы на диапазон [-1..1].

#### Выводы:

1. Кластеризация регионов с учетом главных компонент позволила выделить 5 кластеров, четко различающихся по соотношению затрат, обеспеченности кадрами и здоровья населения.
2. Корреляционный анализ совокупных показателей здоровья и обеспеченности кадрами с показателем затрат на 1 жителя выявил гораздо большую взаимосвязь здоровья населения с обеспеченностью кадрами

- (коэффициент корреляции 0.75) по сравнению с уровнем затрат (коэффициент корреляции 0.38).
3. Полученные результаты позволяют планировать дальнейшее изучение типичных групп регионов с целью более детального выяснения причин, приведших к наблюдающейся ситуации и прицельно искать методы повышения эффективности управления системой здравоохранения в Красноярском крае.
  4. Используемая в работе технология, включающая комбинацию двух математических методов, может быть применена для анализа систем здравоохранения в крупных регионах России, а также, возможно, для анализа функционирования различных объектов системы здравоохранения (ЛПУ, город).

## IV.2. ТАБЛИЦА РЕЗУЛЬТАТОВ ВЫБОРОВ ПРЕЗИДЕНТОВ США

Проиллюстрируем процесс моделирования данных с пробелами на основе таблицы результатов выборов президентов США, которая содержит 31-у предвыборную ситуацию (с 1860 по 1980 гг.), а также одну тестовую ситуацию (1992 г.). Для каждого выбора в таблице содержатся данные по 12-ти бинарным признакам [77], которые перечислены ниже.

1. Правящая партия была у власти более одного срока? (*More1*)
2. Правящая партия получила более 50% голосов на прошлых выборах? (*More50*)
3. В год выборов была активна третья партия? (*Third*)
4. Была серьезная конкуренция при выдвижении кандидата правящей партии? (*Conc*)
5. Кандидат от правящей партии был президентом в год выборов? (*Prez*)
6. Был ли год выборов временем спада или депрессии? (*Depr*)
7. Был ли рост среднего национального валового продукта на душу населения более 2.1%? (*Val2.1*)
8. Произвел ли правящий президент существенные изменения в политике? (*Chan*)
9. Во время правления были существенные социальные волнения? (*Wave*)
10. Администрация правящей партии виновна в серьезной ошибке или скандале? (*Mist*)
11. Кандидат от правящей партии национальный герой? (*R.Hero*)
12. Кандидат от оппозиционной партии национальный герой? (*O.Hero*)

Также в таблице содержится информация о результатах выборов (победе правящей или оппозиционной партии). Значения входных бинарных признаков равны 0 (ответ "Нет") и 1 (ответ "Да"). Значение выходного признака равно 1 (победа правящей партии) и 2 (победа оппозиции).

Построенные по этой таблице модели уверенно предсказывали результаты вторых выборов Рейгана, победу Буша над Дукакисом, обе победы Клинтона [66]. Причем для уверенного разделения на два класса достаточно одной главной кривой. Проиллюстрируем это на следующем примере.

Построим первые две главные кривые (на основе SOC) и заметим, что они образуют пространство, в котором каждая точка имеет свои координаты ( $x$  и  $y$ ). Используя эти координаты, каждую точку можно поместить на плоскость (с обычными декартовыми координатами) – получаем визуализацию исходного множества данных (рис. 4.3).

На этом рисунке квадратики соответствуют победе правящей партии, а кружочки – победе оппозиции. Треугольник соответствует пропущенному значению (победа оппозиции). Видно, что исходного облако данных четко разбивается на две части, одна из которых соответствует победе правящей партии, а другая – победе оппозиции. Поэтому для уверенного предсказания достаточно одного нелинейного фактора (в линейном случае это не так).

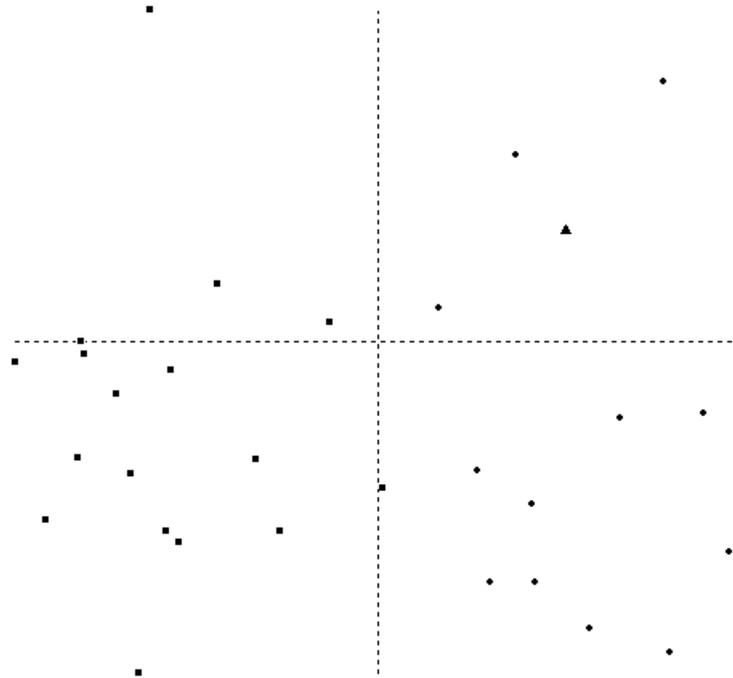


Рис. 4.3. Результаты выборов в пространстве первых двух главных кривых (SOC). Квадратики – победа правящей партии, кружочки – победа оппозиции. Треугольник – отсутствующее значение.

Аналогичный результат получается и другим способом. Если закрыть известные значения результатов выборов и спрогнозировать их значения исходя из модели, то для каждой строчки таблицы получим некоторую ошибку предсказания. Для первой главной кривой распределение этой ошибки представлено в таблице 4.2, из которой видно, что ошибка предсказания не превышает 0.18, что более чем достаточно для уверенного прогноза.

Таблица 4.2.

Интервал	Частота	Проценты
[0.0000, 0.0342)	28 из 31	90.32 %
[0.0342, 0.0683)	1 из 31	3.23 %
[0.0683, 0.1025)	1 из 31	3.23 %
[0.1025, 0.1367)	0 из 31	0.00 %
[0.1367, 0.1708]	1 из 31	3.23 %

Следующим этапом сравнивались степени приближения (в %) исходных данных в зависимости от числа используемых факторов и типов моделей. Средние результаты представлены в таблице 4.3.

Таблица 4.3.

№	Признак	Степень приближения (%) в зависимости от числа факторов								
		Линейная модель			Квазилинейная			SOC		
		1	4	10	1	4	10	1	4	10
1.	Ответ	69.22	69.96	81.78	92.27	92.85	96.72	97.82	98.43	99.50
2.	<i>More1</i>	11.88	59.98	77.36	25.37	63.75	95.91	53.49	80.81	96.85
3.	<i>More50</i>	9.07	61.10	79.43	14.99	73.69	95.18	30.77	75.89	95.12
4.	<i>Third</i>	29.66	44.89	91.56	31.73	66.97	97.45	32.94	76.83	96.93
5.	<i>Conc</i>	62.30	63.28	77.84	69.51	77.12	90.24	72.63	78.86	95.42
6.	<i>Prez</i>	31.72	59.68	80.27	45.58	68.01	93.03	56.08	74.85	95.18
7.	<i>Depr</i>	32.17	52.43	93.38	37.95	71.08	95.56	58.86	80.53	95.62
8.	<i>Val2_1</i>	4.12	37.67	94.19	6.27	69.22	96.53	28.80	72.23	95.44
9.	<i>Changes</i>	2.33	49.87	86.19	16.81	61.15	94.95	13.01	72.01	93.77
10.	<i>Wave</i>	25.13	62.33	80.34	33.18	66.82	95.65	32.68	63.96	96.71
11.	<i>Mist</i>	50.34	61.05	86.17	64.83	70.52	97.55	60.80	81.35	96.90
12.	<i>R_Hero</i>	33.35	48.12	90.69	54.86	66.27	97.52	27.30	83.67	95.76
13.	<i>O_Hero</i>	36.55	50.07	92.03	45.69	68.41	97.55	52.22	76.42	96.17

Если ошибка в вычисленном значении признака меньше 50%, то это означает, что его точное значение (признаки качественные, поэтому при ошибке, меньшей 50%, знак предсказания определяет точное значение).

Для удовлетворительного предсказания линейными моделями достаточно 4-факторов, что говорит о том, что эта задача четырехфакторная (в обычном понимании этого слова). Квазилинейным же моделям, а особенно моделям на основе SOC для удовлетворительного предсказания, как уже отмечалось выше, достаточно всего одного фактора.

Далее проводилось тестирования полученных наборов факторов, которое можно представить в виде двух основных процедур.

1. *Тестирование факторов.* По полной таблице строилась модель, далее в таблицу случайным образом вносились пропуски, а затем запускалась процедура заполнения пропусков. В результате сравнивались получившиеся значения с исходными.

Тестирование показало удовлетворительное заполнения пропусков линейными и квазилинейными моделями при их количестве до 25% от общего числа исходных данных. Тот же показатель для моделей на основе SOC равен 50%. То есть из таблицы можно выкинуть каждое второе данное и все равно их можно будет с удовлетворительной точностью восстановить. Средние значения числа правильно предсказанных примеров для каждого из признаков представлены в таблице 4.4.

Таблица 4.4.

№	Признак	Число правильно предсказанных	Процент
1.	Ответ	11 из 14	78.6 %
2.	<i>More1</i>	10 из 17	58.8 %
3.	<i>More50</i>	13 из 16	81.2 %
4.	<i>Third</i>	9 из 16	56.2 %
5.	<i>Conc</i>	11 из 13	84.6 %
6.	<i>Prez</i>	9 из 12	75 %
7.	<i>Depr</i>	14 из 18	77.7 %
8.	<i>Val2 1</i>	12 из 21	57.1 %
9.	<i>Changes</i>	11 из 17	64.7 %
10.	<i>Wave</i>	10 из 17	58.8 %
11.	<i>Mist</i>	11 из 15	73.3 %
12.	<i>R Hero</i>	12 из 16	75 %
13.	<i>O Hero</i>	12 из 15	80 %

2. *Тестирование метода.* В таблицу случайным образом вносились пропуски, далее по "дырявой" таблице строилась модель, а затем запускалась процедура заполнения пропусков. В результате сравнивались получившиеся значения с исходными.

Удовлетворительное заполнение наблюдалось при количестве пробелов до 10% от общего числа исходных данных.

### **IV.3. ВЕРИФИКАЦИЯ СВЯЗЕЙ МЕЖДУ ДВУМЯ ДИНАМИЧЕСКИМИ СИСТЕМАМИ**

Как указано в [16], проблема верификации связей между двумя динамическими системами представляет особый интерес в гелиофизике. Для этого проводились эксперименты с различными временными рядами. Результаты одного из них приведены на рис. 4.4.

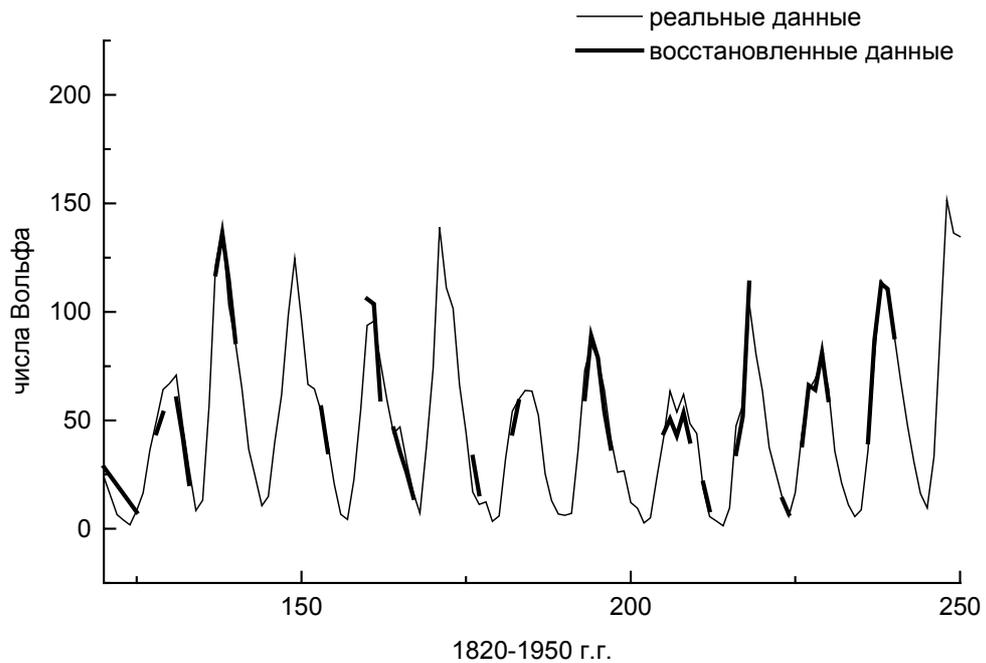


Рис. 4.4. Фрагмент годовых значений ряда Вольфа. SOC, число узлов – 10

Из полного ряда годовых значений чисел Вольфа (для солнечной активности) было удалено 50% точек. Для восстановления пропусков использовалась модель *СОК*. Строки исходной таблицы представляли собой  $m$ -мерные запаздывающие векторы Такенса, с  $m=6$ , вида:  $a_{kj} = x_k, x_{k+1}, \dots, x_{k+5}$ , так что удаление одного отсчета в таблице приводит к удалению соответствующей диагонали. Нейронный конвейер удовлетворительно восстановил даже вершины циклов. На рис. 4.5 приведен фрагмент временного ряда космогенного изотопа  $^{14}\text{C}$ . Полный ряд имеет диапазон с 5995 г.ВС по 1945г.АД. Мы даем результаты восстановления 30% удаленных точек фрагмента временного ряда датированного периодом с 5995 г. ВС по 10 г. АД. Результаты восстановления очень хорошие.

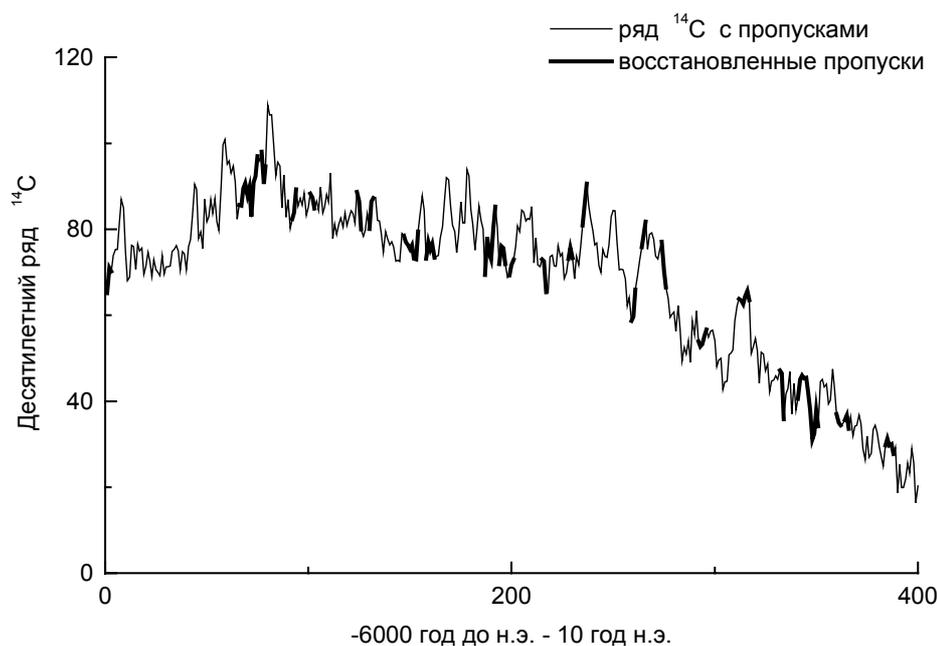


Рис. 4.5. Фрагмент десятилетних значений ряда  $^{14}\text{C}$ .  
Квазилинейная модель, число узлов – 8

Заметим, что составление исходной таблицы по Такенсу, в отличие от произвольного способа приведенного в [1, 14, 78], существенно меняет ситуацию. Действительно, пропущенное значение  $y$ -компоненты одного из векторов ( $L_1$ ) на рис. 4.6 индуцирует пропуск  $x$ -компоненты в следующем (по Такенсу) векторе ( $L_3$ ).

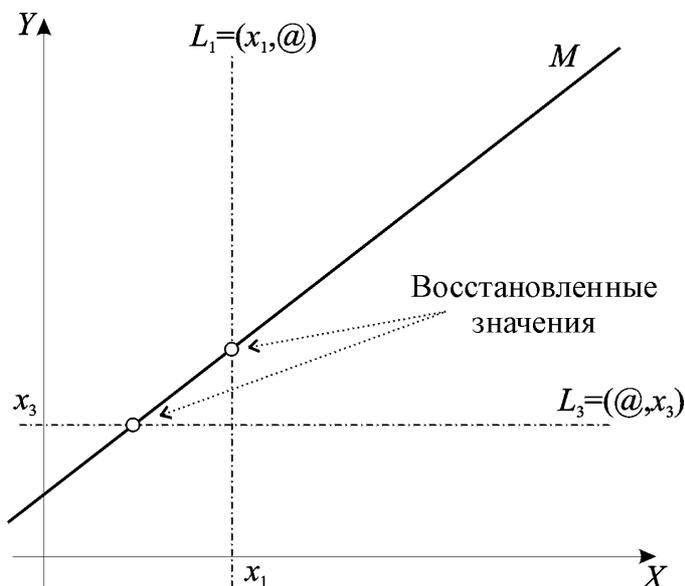


Рис. 4.6.

Пересечение 2-х прямых однозначно восстанавливает пропущенное значение. Для многомерного случая задача сводится к поиску *трансверсального* [45] пересечения плоскостей. Таким образом, метод восстановления пропусков получает формальный контекст.

Наши эксперименты показали, что метод заполнения пробелов с использованием нейронных сетей вполне приемлем для работы с реальными рядами.

#### **IV.4. СОЧЕТАННЫЕ ПОРАЖЕНИЯ ПРОВОДЯЩЕЙ СИСТЕМЫ СЕРДЦА**

В клинике главного кардиолога Красноярского края профессора Шульмана Владимира Федоровича программный продукт ModelAnalyzer нашел применение для выяснения прогноза у пациентов со сложными нарушениями ритма и проводимости сердца [46]. Под наблюдением сотрудников этой клиники в течении 15-20 лет находились пациенты с соответствующей патологией. Сотрудники клиники ежегодно проводят обследования данной группы пациентов, выясняя клиническое течение заболеваний, частоту обострения данной патологии, качество жизни данной группы пациентов. При столь длительном наблюдении постоянно накапливаются данные для выяснения, какие же факторы в большей степени влияют на прогноз данной патологии.

Существенную помощь в решении этого вопроса оказала методика многомерного факторного анализа. С ее использованием было выяснено, что у больных с сочетанными поражениями проводящей системы сердца в сравнении с пациентами с изолированным поражением проводящей системы сердца значительно чаще наблюдается застойная сердечная недостаточность, тромбоэмболические осложнения, некоторые формы нарушения сердечного ритма.

Кроме того были выявлены факторы, способствующие возникновению ряда осложнений. В частности, возникновению и развитию застойной сердечной недостаточности способствует наличие пароксизмальной формы фибрилляции предсердий и органические заболевания сердца в анамнезе. Возникновению тромбоэмболии способствует наличие фибрилляции и трепетания предсердия в анамнезе пациента. Возникновению фибрилляций предсердий способствует застойная сердечная недостаточность и нарушение прироста частоты сердечных сокращений при физической нагрузке.

Методика многомерного факторного анализа помогла выяснить, что у пациентов со сложными нарушениями ритма не эффективны некоторые режимы постоянной электрокардиостимуляции, что безусловно послужит темой дальнейших исследований в этой области. И, наконец, были выявлены факторы, способствующие возникновению летального исхода: у больных с сочетанными поражениями проводящей системы сердца наиболее неблагоприятно для прогноза наличие тромбоэмболий, застойной сердечной недостаточности в анамнезе, а также возникновение постоянной или пароксизмальной формы фибрилляции предсердий.

Таким образом, методика многомерного факторного анализа может с успехом использоваться для выявления факторов, оказывающих влияние на

течение и прогноз заболевания у больных со сложными нарушениями ритма и проводимости сердца.