

Факторный анализ результатов вейвлет-преобразований лонгитюдных данных как новый метод исследования динамических характеристик сложных систем

Представлен новый метод выявления и исследования факторов, определяющих развитие сложных систем. Предложенный подход опирается на возможности вейвлет-преобразований и идентифицируемых факторных структур. В соответствии с ним, значения коэффициентов, полученные в результате дискретного вейвлет-преобразования временного ряда наблюдаемого процесса и соответствующие различным периодам наблюдений, рассматриваются как значения наблюдаемых переменных в последующем конфирматорном факторном анализе, который, в свою очередь, используется для выявления динамики факторных влияний и оценки показателей взаимодействий между факторами. Идентификация свободных параметров факторной модели (как правило, факторных дисперсий и ковариаций) выполняется с помощью новой прямой (неитерационной) процедуры, опирающейся на метод максимального правдоподобия, что является альтернативой традиционному итерационному поиску локального решения задачи многомерной численной оптимизации. Рассматривается метод оценки статистической значимости компонентов факторной модели. Представлены преимущества данного подхода перед традиционным симплекс-методом, а также ряд способов построения факторных моделей, представленных путевыми диаграммами, включая их сравнительный анализ.

1. ВВЕДЕНИЕ

Как правило, доступные для измерения параметры, исследуемые при мониторинге системы, не представляют её характеристики в форме, удобной для непосредственного выявления текущего состояния и формулирования надёжных условий, достаточных для надлежащей диагностики. В случае многомерных измерений, с которыми обычно имеют дело при мониторинге, важно выявить несколько латентных факторов, отвечающих за изменчивость наблюдаемых параметров, доступных для измерения, и определить их природу и диапазон воздействий, используя в дальнейшем полученную информацию для идентификации состояния системы.

При этом, обеспечивая минимальные потери полезной информации, параметры, которые легко измерять, следует, по возможности, заменять на параметры, которые легко интерпретировать и использовать в дальнейшем для исследования поведения системы. При последующем анализе должны быть определены функциональные зависимости наблюдаемых параметров от выявляемых факторов. В результате должна быть выявлена вся структура причинных связей между факторами и наблюдаемыми переменными, а также, если потребуется, непосредственные значения факторов, необходимые для идентификации состояния системы.

Для того, чтобы выполнить приведенные выше требования, были разработаны эмпирические математические модели и соответствующие методы многомерного статистического анализа [2-3, 5, 11-12]. Наиболее приемлемыми в рассматриваемой ситуации являются исследовательские (эксплораторные) и проверочные (конфирматорные) факторные модели и методы их исследования. Оба подхода работают с выборочными матрицами ковариаций или корреляций наблюдаемых переменных. Исследовательский анализ предполагает наличие некоторого неизвестного заранее числа некоррелированных факторов с неопределенной интерпретацией, тогда как при конфирматорном факторном анализе факторы, их интерпретация и причинные связи с наблюдаемыми переменными, а также корреляционные связи между латентными факторами определяются постановкой задачи. Конфирматорные модели позволяют с помощью несложной процедуры оценивать статистическую значимость каждого своего компонента.

Поскольку в задачах, наиболее часто встречающихся на практике, обычно известны гипотезы о причинах возможных влияний на наблюдаемые переменные, то последний подход, как правило, предпочтительней.

В задачах диагностического мониторинга в большинстве случаев исследуется временная динамика наблюдаемых переменных, которые в различных контрольных точках формально рассматриваются как отдельно анализируемые величины. Для выявления статистических связей между влияющими на них факторами разработан симплекс-метод конфирматорного факторного анализа [6]. Однако практическое применение этого подхода обусловлено рядом серьезных ограничений, которые зачастую делают его использование невозможным. В частности, для анализа пригодны только ковариационные и корреляционные матрицы с симплексной структурой; факторное взаимодействие можно исследовать только для смежных контрольных точек, при этом невозможно связать факторы с периодами времени, к которым они относятся, и т.д. Кроме того, существенным недостатком традиционного конфирматорного факторного анализа является необходимость решения трудоемкой задачи многомерной локальной оптимизации для оценки величин свободных параметров модели, что, как правило, не позволяет найти глобальный минимум и приводит к неоднозначности решения.

Чтобы преодолеть указанные проблемы, был разработан представленный далее новый подход [7], опирающийся на возможности вейвлет-преобразований и идентифицируемых факторных структур. К его особенностям и преимуществам относятся:

- возможность нахождения оценок свободных параметров модели прямыми (неитерационными) методами, гарантирующими однозначное оптимальное решение;
- гибкие средства для исследования взаимодействия факторов;
- применимость для анализа ковариационных и корреляционных матриц произвольной структуры.

Рассмотренные далее идентифицируемые факторные структуры фактически выступают в качестве одной из разновидностей нейронных сетей. Элементы таких сетей соответствуют латентным и наблюдаемым переменным факторных моделей и являются аналогами традиционных нейронов. Нелинейное преобразование взвешенной суммы значений на входе в факторных моделях не производится, поэтому функции активации нейрона соответствует тождественная функция. В качестве весов используются факторные нагрузки, а входами и выходами сетевых элементов являются значения факторов и наблюдаемых переменных. Более подробно список аналогий представлен в таблице 1.

Таблица 1. Аналогии между нейронными сетями и факторными структурами.

Нейронная сеть	Факторная структура
Нейрон	Фактор (латентная переменная) или наблюдаемая переменная
Функция активации нейрона	Тождественная функция (нелинейное преобразование взвешенной суммы значений на входе не производится)
Веса соединений	Факторные нагрузки
Порог активации	—
Входы и выходы нейронов	Значения факторов и наблюдаемых переменных
Свободные параметры есть веса соединений и пороги активации	Свободные параметры есть факторные нагрузки и ковариации (корреляции) между факторами
Обучающая выборка	Выборочные оценки матриц ковариаций, полученные в результате наблюдений

При обучении минимизируется ошибка прогноза сети в заданной метрике	При обучении минимизируется статистика, асимптотически описываемая распределением χ^2 и определяющая различия между наблюдаемой и прогнозируемой матрицами ковариаций
---	--

2. ОСНОВНЫЕ ЭТАПЫ АНАЛИЗА

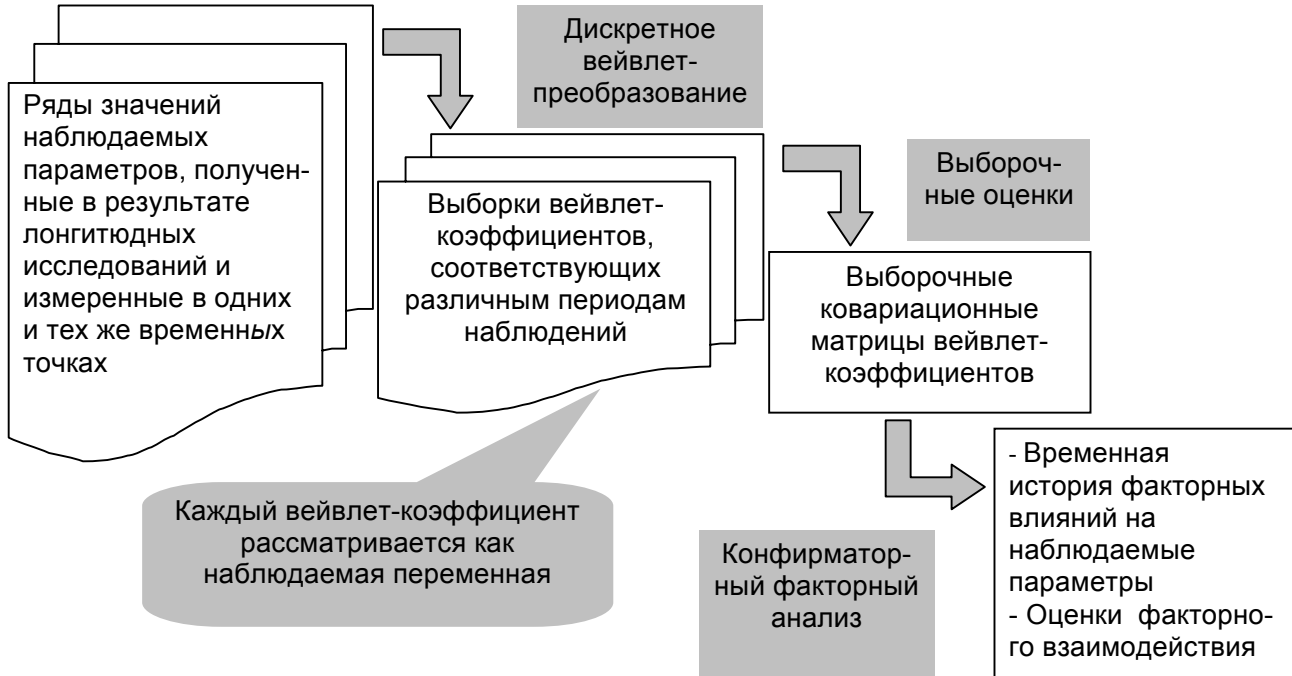


Рис. 1. Основные этапы анализа.

Основные этапы предложенного подхода представлены на рис. 1. Выборки коэффициентов, полученных в результате дискретных вейвлет-преобразований временных рядов значений исследуемого параметра и соответствующих различным периодам наблюдений, в последующем конфирматорном факторном анализе рассматриваются как значения наблюдаемых переменных, которые позволяют выявить временную историю факторных влияний и оценки факторного взаимодействия. Представление данных, полученное с помощью вейвлет-преобразований, даёт возможность выявить различия в характеристиках процесса в различных шкалах. Идентификация свободных параметров факторной модели (обычно корреляций или ковариаций) выполняется с помощью новой прямой (неитерационной) процедуры, которая опирается на метод максимального правдоподобия и является альтернативой традиционному итерационному способу решения задач локальной оптимизации.

2.1. Вейвлет-преобразования

Рабочие представления анализируемых процессов формируются с помощью вейвлет-преобразования, которое позволяет выявлять различия в их характеристиках при различных шкалах измерений на всём протяжении интервала наблюдений. Если исследуемый процесс есть функция одной переменной, то его вейвлет-спектр – функция двух аргументов, один из которых характеризует период составляющих компонентов, а другой – смещение вычисляемых показателей вдоль оси времени. Для вычисления вейвлет-спектров используются вейвлеты – особые функции в форме коротких волн («всплесков») с нулевым интегральным значением и локализацией по оси независимой переменной, способные к сдвигу по этой оси и масштабированию.

Вейвлет-анализ имеет очевидные преимущества перед традиционным спектральным анализом, поскольку он обеспечивает корректные результаты в случае нестационарных

процессов и содержит более полную информацию о поведении изучаемого объекта. Это сделало данный подход популярным среди исследователей разных специальностей. Далее применялся дискретный вариант этого метода, позволивший представить анализируемые процессы в виде элементов определённого метрического функционального пространства с вейвлет-базисом.

2.2. Альтернативный вариант конфирматорного факторного анализа

Основные компоненты конфирматорного факторного анализа представлены на рис 2.



Рис. 2. Основные компоненты конфирматорного факторного анализа.

Примечание. В случае нескольких типов наблюдаемых случайных процессов для вейвлет-коэффициентов формируется смешанная ковариационная матрица.

2.2.1. Традиционный конфирматорный факторный анализ

Конфирматорный факторный анализ предполагает наличие строго определённой факторной модели изучаемого явления. Факторная модель, связывающая латентные и наблюдаемые переменные, формируется, опираясь на знание предметной области. Гипотезы о структуре модели должны основываться на анализе природы исследуемых факторов. Можно делать количественные предположения о значениях факторных нагрузок и корреляциях между факторами, а также проверять гипотезы о структуре и свойствах моделей, подбирая их оптимальные варианты [13]. Параметры моделей подбираются так, чтобы обеспечить наилучшее с точки зрения заданного критерия приближение к ковариационным (корреляционным) матрицам наблюдаемых переменных.

Объектами данного вида анализа являются ковариационные или корреляционные матрицы наблюдаемых переменных. Цель работы – выявить значения параметров модели, которая с приемлемыми ошибками объясняет изменчивость наблюдений.

При использовании этого подхода:

- ненулевые (свободные) факторные нагрузки и число исследуемых факторов в модели определяются заранее;
- допускаются корреляции между ошибками измерений;
- факторные нагрузки и ковариации между латентными переменными могут быть свободными параметрами модели или приравняться заданным константам;
- допускается анализ нескольких групп моделей;
- можно проверять, насколько согласуются ограничения, налагаемые на параметры модели, с результатами наблюдений.

Для определения оценок свободных параметров модели методом максимального правдоподобия в качестве минимизируемого критерия используется функция

$$F = [\ln|\Sigma| - \ln|S| + \text{tr}(S\Sigma^{-1}) - q] (N-I),$$

где S – выборочная ковариационная матрица наблюдаемых переменных, Σ – прогнозируемая ковариационная матрица наблюдаемых переменных, $|\Sigma|$ и $|S|$ – определители матриц Σ и S , $\text{tr}(S\Sigma^{-1})$ – след матрицы $(S\Sigma^{-1})$, N – объем выборки, использованной для вычисления матрицы S , q – число наблюдаемых переменных [11].

Элементы прогнозируемой ковариационной матрицы представляют собой аналитические выражения относительно свободных параметров модели. В случае многомерного нормального распределения наблюдаемых переменных значения критерия F описываются распределением χ^2 .

Для определения свободных параметров модели необходимо численно решить итерационными методами достаточно трудоемкую задачу локальной многомерной оптимизации. Из этого в общем случае вытекает невозможность определения глобального минимума, так как в результате решения находится один из локальных минимумов, выбор которого зависит от начального приближения. Таким образом, решение неоднозначно.

2.2.2. Альтернативный вариант конфирматорного факторного анализа: основные принципы подхода

Для определения свободных параметров модели традиционный конфирматорный факторный анализ предполагает решение трудоёмкой задачи многомерной локальной оптимизации. Это приводит к невозможности определения глобального минимума и неоднозначности решения. Предлагаемый альтернативный вариант конфирматорного факторного анализа позволяет находить единственное оптимальное решение прямыми (неитерационными) методами. Процедура решения включает в себя следующие этапы:

- составление переопределенной системы алгебраических уравнений, выражая выборочные дисперсии и ковариации через аналогичные факторные показатели;
- решение полученной системы прямым (неитерационным) методом, используя метод максимального правдоподобия [8-9];
- проверку адекватности полученных моделей наблюдениям, опираясь на статистические критерии согласия.

Чтобы избежать составления сложных для решения нелинейных систем уравнений относительно коэффициентов корреляции и факторных нагрузок, используется путевая модель дисперсионных составляющих [13], в которой факторные нагрузки равны единице. Каждой наблюдаемой дисперсии и ковариации ставится в соответствие алгебраическое уравнение, которое связывает её выборочную оценку с соответствующей прогнозируемой величиной, выраженной аналитически через неопределённые дисперсии и ковариации латентных переменных [3, 12]. В путевых диаграммах для этого, в частности, могут быть использованы правила обхода путей. В результате получается система, число уравнений которой равно числу наблюдаемых дисперсий и ковариаций. Для вычисления оценок максимального правдоподобия и проверки адекватности модели необходимо, чтобы

значения наблюдаемых переменных описывались многомерным нормальным распределением, а число уравнений в исследуемой системе превышало число свободных параметров модели.

Представим полученную переопределенную систему n уравнений в матричной форме:

$$\mathbf{Ax}=\mathbf{b},$$

где \mathbf{A} – матрица системы, коэффициенты которой определяются факторной моделью; \mathbf{b} – вектор-столбец n выборочных дисперсий и ковариаций, определяемых результатами наблюдений; \mathbf{x} – вектор-столбец m искомым дисперсий и ковариаций латентных переменных.

Теперь рассмотрим вектор $\boldsymbol{\varepsilon}=\mathbf{Ax}-\mathbf{b}$, представляющий полученную методом наименьших квадратов невязку псевдорешения \mathbf{x} переопределенной системы. Полагая в общем случае, что компоненты вектора невязок коррелированы, выразим их невырожденную ковариационную матрицу как $\sigma^2\mathbf{V}$.

Сделаем замены

$$\mathbf{b}=\mathbf{V}^{1/2}\mathbf{b}_0 \text{ и } \mathbf{A}=\mathbf{V}^{1/2}\mathbf{A}_0,$$

где $\mathbf{V}=\mathbf{V}^{1/2}\mathbf{V}^{1/2}$, перейдем к системе:

$$\mathbf{A}_0\mathbf{x}=\mathbf{b}_0,$$

ковариационная матрица вектора невязок $\boldsymbol{\varepsilon}_0=\mathbf{V}^{-1/2}\boldsymbol{\varepsilon}$ которой имеет вид $\sigma^2\mathbf{E}$, где \mathbf{E} – единичная матрица.

Если рассматриваемая система невырождена ($\text{rank}\mathbf{A}=m$, где m – число свободных параметров модели), вектор невязки $\boldsymbol{\varepsilon}_0$ имеет многомерное нормальное распределение, а

$$\mathbf{x}=(\mathbf{A}_0^T\mathbf{A}_0)^{-1}\mathbf{A}_0^T\mathbf{b}_0=(\mathbf{A}^T\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{V}^{-1}\mathbf{b}-$$

псевдорешение, полученное методом наименьших квадратов, то это псевдорешение является оценкой максимального правдоподобия, а статистика

$$X^2=(\mathbf{b}_0-\mathbf{A}_0\mathbf{x})^T(\mathbf{b}_0-\mathbf{A}_0\mathbf{x})/\sigma^2=(\mathbf{b}-\mathbf{Ax})^T\mathbf{V}^{-1}(\mathbf{b}-\mathbf{Ax})/\sigma^2$$

имеет распределение χ^2 с $n-m$ степенями свободы [14].

Указанная статистика X^2 позволяет, при заданных выше предположениях, проверять гипотезу о представимости выборочных дисперсий и ковариаций, составляющих вектор \mathbf{b} , дисперсиями и ковариациями латентных переменных, содержащихся в исследуемой модели.

Область принятия гипотезы есть $X^2 \leq \chi^2_{n-m;\alpha}$, где α есть уровень значимости критерия.

При реализации данного подхода удобно сделать следующие упрощающие предположения, обусловленные особенностями искомого решения:

- компоненты вектора невязок $\boldsymbol{\varepsilon}$ являются некоррелированными;
- значения среднеквадратических отклонений различных компонентов вектора невязок $\boldsymbol{\varepsilon}$ составляют одну и ту же фиксированную долю (процент) от соответствующих компонентов вектора \mathbf{b} .

Чтобы обеспечить сопоставимость оценок, указанная доля (процент) подбирается так, чтобы равенство $X^2=\chi^2_{n-m;\alpha}$ выполнялось при $\alpha=0.05$. Для оценки степени допустимости вычисленной характеристики для неё удобно установить разумное критическое значение, например, 0.1 . Таким образом, вместо уровня значимости появляется новый показатель – критический процент.

Преимуществом предложенного подхода является то, что он не сводит решение к трудоёмкой процедуре многомерной локальной оптимизации, которая не гарантирует нахождение глобальных минимумов, обеспечивая при этом однозначность результата.

Рассмотренный метод решения позволяет выявлять характер взаимных связей между свободными параметрами факторной модели путём прямого расчёта по матричным формулам, численно вычисляя оценки одних характеристик для заданного множества сочетаний значений других.

Для ряда моделей для обеспечения корректного результата на оцениваемые дисперсии и ковариации следует налагать дополнительные условия в форме неравенств, выражающие положительность значений дисперсий и корректность соотношений между связанными дисперсиями и ковариациями. В этом случае в качестве решения принимается вектор \mathbf{X}_c , удовлетворяющий указанным условиям и ближайший к найденному псевдорешению \mathbf{X} в евклидовой метрике. Для поиска вектора \mathbf{X}_c решается несложная задача линейной оптимизации. Гипотеза о представимости выборочных дисперсий и ковариаций, составляющих вектор \mathbf{b} , дисперсиями и ковариациями латентных переменных, составляющими вектор \mathbf{X}_c , проверяется, как и ранее, с помощью статистики

$$X^2 = (\mathbf{b} - \mathbf{A}\mathbf{X}_c)^T \mathbf{V}^{-1} (\mathbf{b} - \mathbf{A}\mathbf{X}_c) / \sigma^2,$$

имеющей распределение χ^2 с $n-m$ степенями свободы.

Как и в традиционном конфирматорном факторном анализе, рассматриваемый метод даёт возможность строить заключения о статистической значимости различных компонентов модели, используя статистические критерии согласия.

Для этого следует сравнить статистики X^2 для двух моделей: полной модели, содержащей исследуемый компонент, и упрощенной модели, в которой этот компонент отсутствует. Гипотезу о том, что полная модель согласуется с результатами наблюдений, будем обозначать как H_f . Выявление степени значимости исследуемого компонента производится, если отвергать гипотезу H_f нет оснований. Сначала следует оценить свободные параметры упрощенной модели. Полученное значение статистики X^2 для упрощенной модели сравнивается с аналогичной характеристикой для полной модели.

Поскольку разность указанных статистик асимптотически распределена как χ^2 с числом степеней свободы, равным разности в числах степеней свободы полной и упрощенной моделей, эта разность используется для проверки нулевой гипотезы H_r о том, что упрощенная модель согласуется с результатами наблюдений, против альтернативной гипотезы H_f .

Если гипотеза H_r не отвергается при заданном уровне значимости, то исследуемый компонент признается статистически незначимым и делается вывод о том, что имеющиеся данные не свидетельствуют о его влиянии на данную характеристику. Если гипотеза H_r отвергается (а гипотеза H_f принимается), то можно говорить о влиянии исследуемого компонента на эту характеристику.

2.3. Традиционный подход к лонгитюдным исследованиям: симплекс-модели

Симплекс модели [6] в лонгитюдных исследованиях обычно используются в случае, когда корреляционные матрицы наблюдаемых параметров имеют симплексную структуру, что подразумевает меньшие корреляции между параметрами при больших временных интервалах между точками измерений. Примеры типичных одно- и двухфакторных симплекс-моделей представлены на рис. 3 и 4 в форме путевых диаграмм, где η_i , G_i и E_i -

факторы, воздействующие в i -ой точке измерений, β_i – факторные нагрузки, ζ_i – случайные воздействия, некоррелированные с η_{i-1} (инновации), y_i – наблюдаемые переменные, ε_i – ошибки измерений.

Связи между факторами в различные моменты времени представлены авторегрессионными зависимостями:

$$\eta_i = \beta_i \eta_{i-1} + \zeta_i, \quad G_i = \beta_{ai} G_{i-1} + \zeta_{ai} \quad \text{и} \quad E_i = \beta_{ei} E_{i-1} + \zeta_{ei}.$$

Модели измерений выражаются как

$$y_i = \lambda_i \eta_i + \varepsilon_i,$$

где λ_i – нормализующие коэффициенты, для однофакторной модели и как

$$y_i = G_i + E_i + \varepsilon_i$$

для двухфакторной модели.

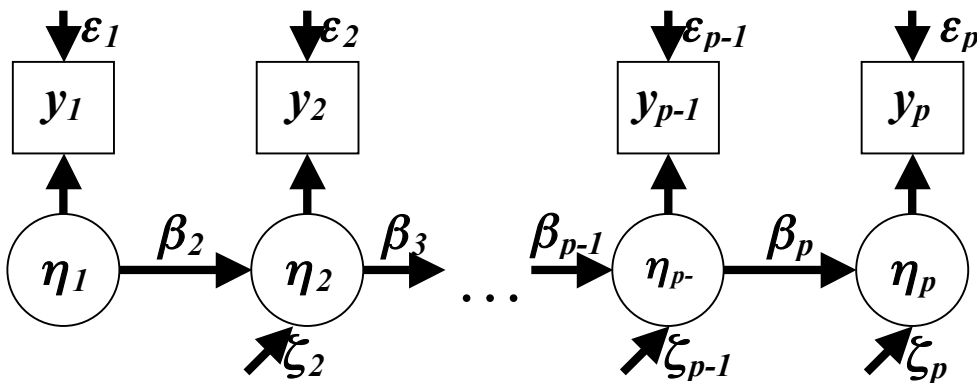


Рис.3. Однофакторная симплекс-модель.

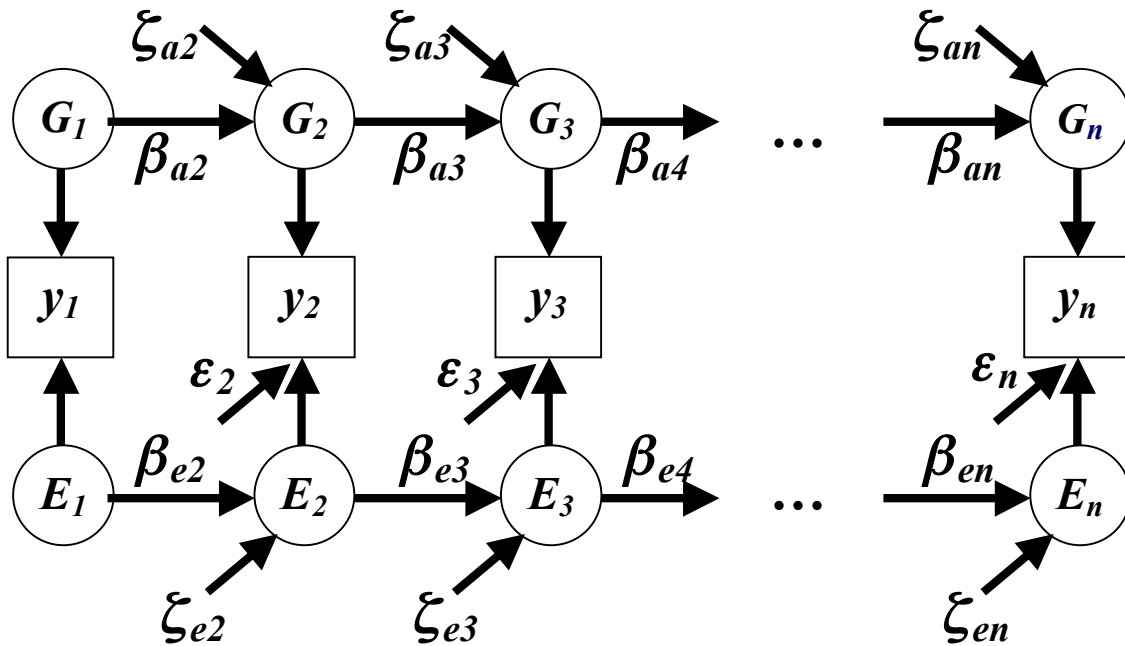


Рис.4. Двухфакторная симплекс-модель.

2.4. Альтернативные факторные модели для анализа результатов вейвлет-преобразований

Для создания новых лонгитюдных факторных моделей, опирающихся на конфирматорный факторный анализ результатов вейвлет-преобразований, применяются модели путевых коэффициентов и дисперсионных составляющих, включая их модификации. Примеры типовых вариантов этих моделей показаны на рис. 5 и 6. Состав вейвлет-коэффициентов, используемых при анализе в качестве наблюдаемых переменных, зависит от рассматриваемой прикладной задачи и может меняться. Обычно предполагается, что число моментов времени, в которые производятся наблюдения, является степенью числа 2.

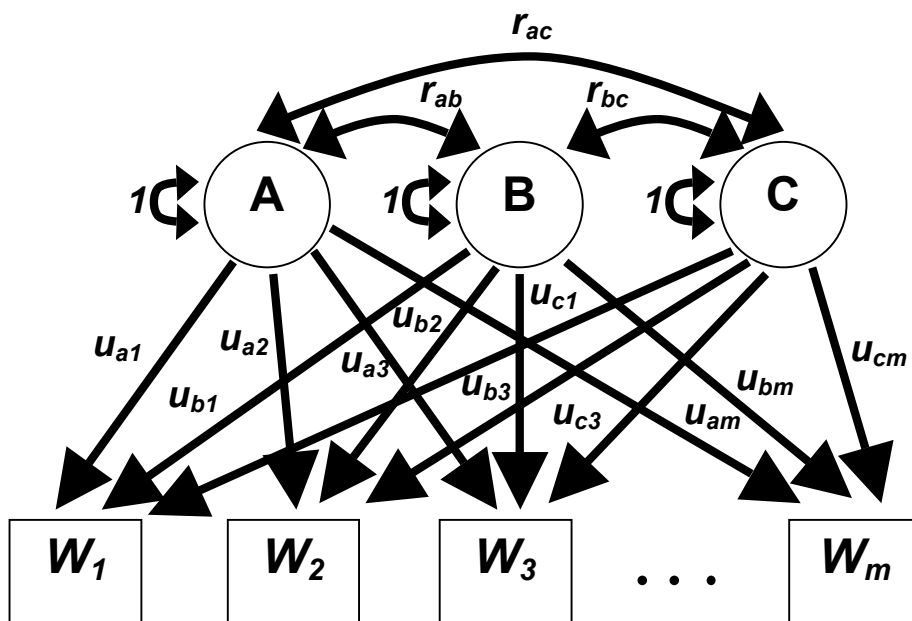


Рис. 5. Модель путевых коэффициентов, представленная путевой диаграммой.

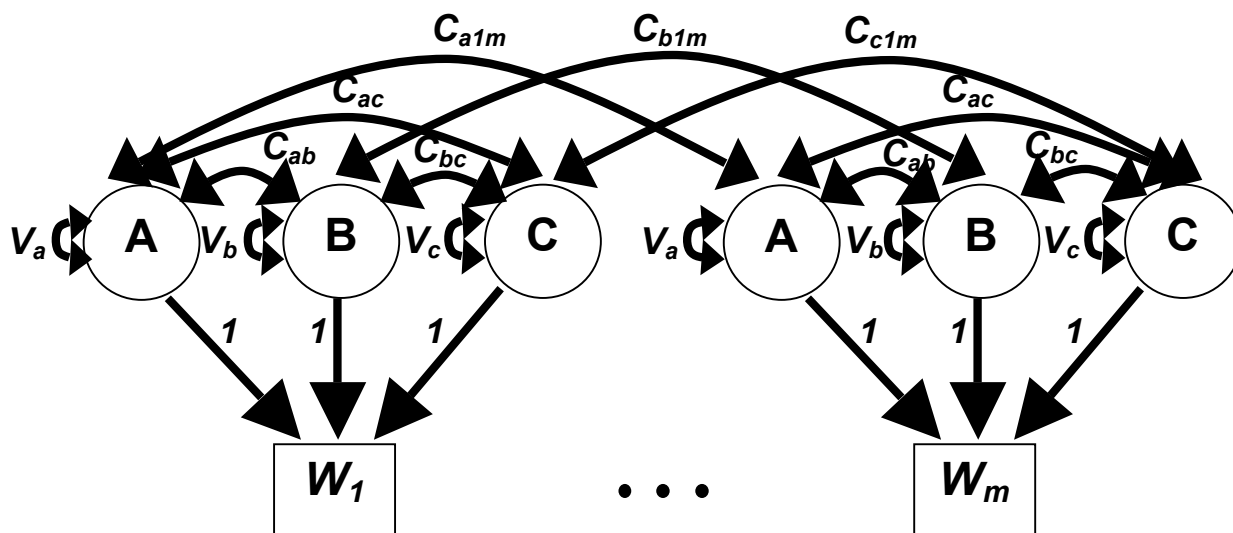


Рис. 6. Модель дисперсионных составляющих, представленная путевой диаграммой.

В случае модели путевых коэффициентов аналитические выражения для ковариаций и дисперсий вейвлет-коэффициентов W_i нелинейны:

$$Cov(W_i, W_j) = \sum_k \sum_l r_{kl} u_{ki} u_{lj},$$

$$Var(W_i) = \sum_k \sum_l r_{kl} u_{ki} u_{li},$$

где k и l – индексы факторов, u_{**} – факторные нагрузки, r_{**} – корреляции между факторами, – что не позволяет получать простые и однозначные оценки свободных параметров модели. Подобные выражения в случае модели дисперсионных составляющих являются линейными:

$$Cov(W_i, W_j) = \sum_k C_{kij},$$

$$Var(W_j) = \sum_k V_k + \sum_k \sum_l C_{kl},$$

где k и l – индексы факторов, V_* – дисперсии, C_{**} и C_{***} – ковариации между факторами, что даёт возможность получать прямые оценки свободных параметров, используя рассмотренный выше альтернативный вариант конфирматорного факторного анализа, и применять данную модель на практике.

Отражая важные для исследования особенности прикладных задач, модель дисперсионных составляющих принимает различные частные формы. Например, при изучении факторных влияний в различных условиях может быть полезен одновременный анализ различных групп моделей (рис. 7).

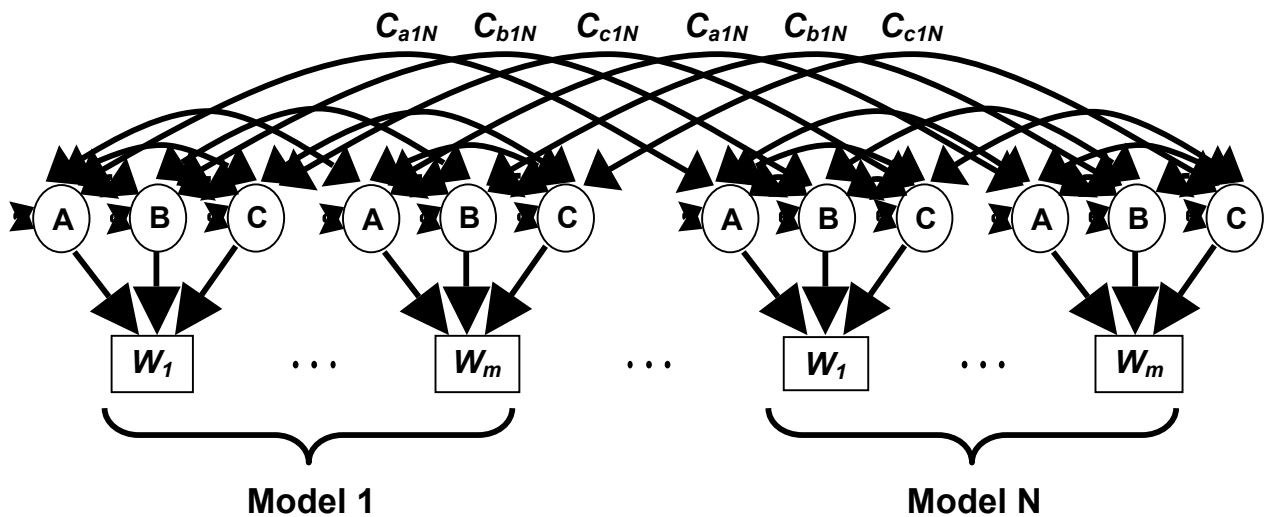


Рис. 7. Изучение факторных влияний в различных условиях: одновременный анализ различных групп моделей. C_{***} – ковариации между факторами.

Интерпретация результатов рассмотренного варианта факторного анализа обычно опирается на:

- оценки свободных факторных дисперсий и ковариаций;
- оценки свободных корреляций между различными факторами в одни и те же моменты времени;
- оценки свободных корреляций между одинаковыми факторами в разные моменты времени;
- оценки статистической значимости различных компонентов модели.

Соответствующие примеры могут быть найдены в статье [1].

Сопоставление преимуществ представленного подхода и недостатков симплекс-метода представлено в таблице 2.

Таблица 2. Сопоставление преимуществ альтернативного варианта конформаторного факторного анализа и недостатков симплекс-метода.

Симплекс-метод: недостатки <i>Назначение:</i> метод позволяет исследовать баланс влияний факторов из прошлого и новых влияний (инноваций)	Альтернативный вариант конформаторного факторного анализа: преимущества <i>Назначение:</i> метод позволяет сопоставлять факторные влияния в различные периоды времени
Менее гибкий	Более гибкий
Высокая степень нелинейности элементов прогнозируемой ковариационной матрицы, приводящая к неоднозначному решению	В случае путевой модели дисперсионных составляющих элементы прогнозируемой ковариационной матрицы представляются линейными выражениями, что обеспечивает единственность решения
Приемлем только для ковариационных и корреляционных матриц с симплексной структурой	Подходит для произвольных ковариационных и корреляционных матриц
Меньший запас свободных параметров модели (меньшая степень переопределенности)	Большой запас свободных параметров модели (большая степень переопределенности)
Нельзя связывать факторы с периодами времени	Можно связывать факторы с периодами времени
При большом числе контрольных точек многофакторные модели трудно поддаются анализу	При большом числе контрольных точек многофакторные модели не вызывают проблем при анализе
Возможности для исследования взаимодействия факторов ограничены (только для смежных контрольных точек)	Гибкие возможности для исследования взаимодействия факторов
В модели должны быть учтены все существующие ковариационные связи	Учитывать все существующие ковариационные связи в модели не обязательно (следует рассматривать только те связи, которые важны для проводимого анализа)

2.5. Степень переопределённости модели

Чтобы оценить степень соответствия факторной модели накопленным результатам наблюдений, вычисляется статистика χ^2 , значения которой описываются распределением χ^2 . Число наблюдаемых статистик должно превышать число свободных параметров модели, причём разность между этими показателями равна числу степеней свободы указанного распределения. Показывая число дополнительных свободных параметров, которые могут быть включены в рассматриваемую модель, число степеней свободы выражает степень переопределенности модели и является важной характеристикой возможностей её практического применения. Выражения этой величины через длины временных реализаций исследуемых параметров и числа факторов для различных типов моделей показаны в таблице 3.

Таблица 3. Числа степеней свободы для различных типов моделей ($M=2^n$ – длина временной реализации исследуемого параметра, K – число факторов)

Тип модели	Число степеней свободы
Симплекс-модель	$M(M+1)/2 - K(3M-2)$
Модель путевых коэффициентов	$M(M+1)/2 - K(M + K - 1)$
Модель дисперсионных составляющих	$M(M+1)/2 - (K^2+K)\log_2 M/2 - K(\log_2 M-1)$ (в случае $\log_2 M-1$ факторных ковариаций для пар вейвлет-коэффициентов)

Наилучшими на практике обычно оказываются модели дисперсионных составляющих. Для иллюстрации этого факта на рис. 8 представлены числа степеней свободы трёхфакторных моделей как функции длин временных реализаций.

Если это целесообразно для рационального решения задачи, то в модели могут быть использованы только некоторые из вейвлет-коэффициентов,.

Для обеспечения переопределенности модели в случае недостаточного числа независимых наблюдаемых статистик следует сократить число свободных параметров, приравнявая их друг к другу, если это допускается имеющимися результатами наблюдений (см. пример на рис. 9).

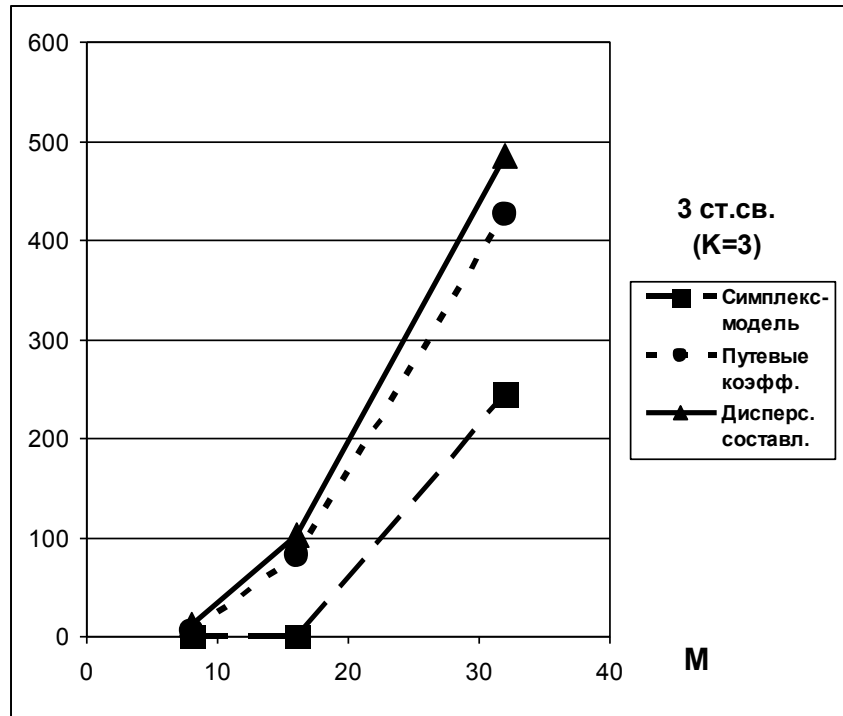


Рис. 8. Степень переопределённости (число степеней свободы) трехфакторной модели ($K=3$) как функция длин временных реализаций (M).

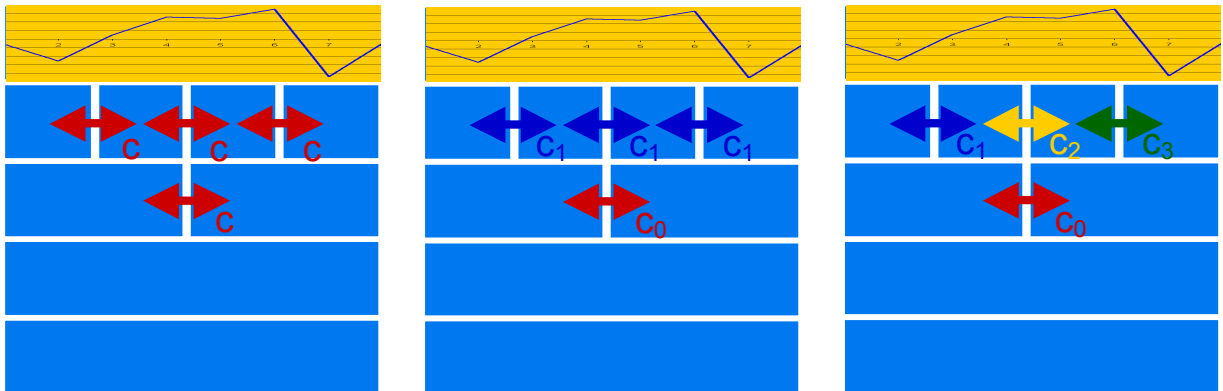


Рис. 9. Примеры различных способов уменьшения числа свободных ковариаций в модели для обеспечения её переопределённости в случае 8-точечной временной реализации исследуемого параметра.

2.6. Сингулярность модели

Если модель, построенная для решения прикладной задачи, приводит к матрице системы линейных уравнений \mathbf{A} , ранг которой меньше, чем число свободных параметров, то псевдорешение $\mathbf{x} = (\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{V}^{-1} \mathbf{b}$ не может быть вычислено однозначно из-за вырожденности матрицы $\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A}$. В этом случае следует уменьшить число свободных параметров модели, исключив зависимые, и обеспечить таким образом невырожденность указанной матрицы. Число подлежащих сокращению параметров равно дефекту матрицы $\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A}$. Для их определения предлагается следующий приём:

- решить алгебраическую проблему собственных значений для матрицы $\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A}$, которая по построению симметрична и неотрицательно определена, установив

подпространство, заданное собственными векторами, соответствующими ненулевым собственным значениям данной матрицы;

- провести вращение базиса полученного собственного подпространства, не выходя за его пределы, с целью достижения наибольшего соответствия между направлениями координатных осей собственного подпространства и осей исходного полного базиса, что формально приведёт к преобразованию координат базисных векторов в исходном пространстве либо в относительно большие, либо в относительно малые значения (рис. 10).

Оси исходного базиса, которые во всех выражениях для повернутых базисных векторов собственного подпространства представлены относительно малыми координатными значениями, задают направления, почти ортогональные вычисленному собственному подпространству с ненулевыми собственными значениями. Таким образом, эти направления приближённо определяют подпространство, соответствующее нулевым собственным значениям, и, следовательно, указывают на зависимые параметры, которые следует исключить из модели для обеспечения невырожденности матрицы. Исключение таких параметров путём выражения их через независимые характеристики или присваивания постоянных значений обычно приводит к устранению дефекта матрицы $\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A}$.

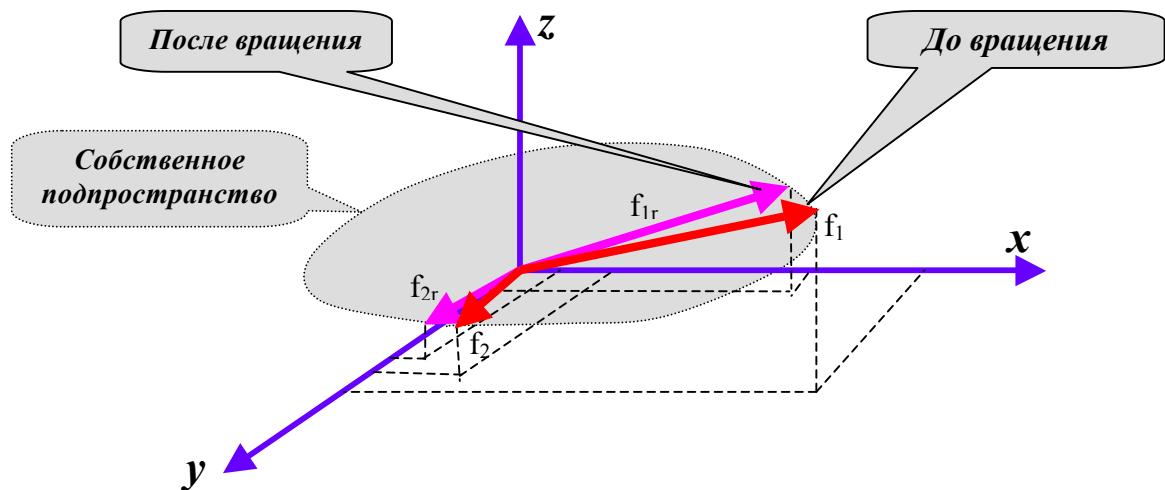


Рис 10. Пример: вращение базиса (\mathbf{f}_1 , \mathbf{f}_2) собственного двухмерного подпространства с целью получения наибольшего соответствия между новыми направлениями координатных осей (\mathbf{f}_{1r} , \mathbf{f}_{2r}) собственного подпространства и осями исходного трехмерного базиса.

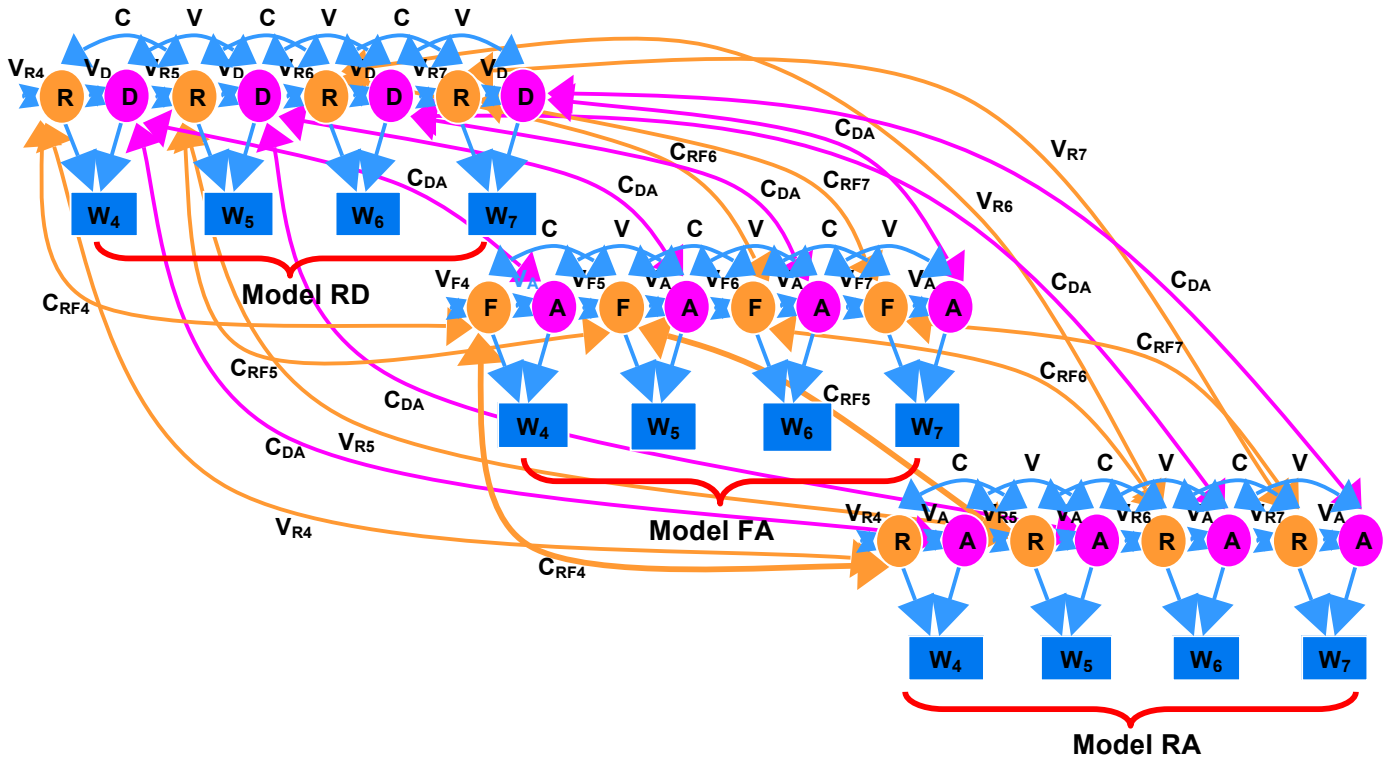
Если эти преобразования приводят к явно неприемлемой модели, то можно, сохранив её первоначальное представление, вычислить псевдорешение приближённо, используя итерационный метод Гаусса-Зейделя или другие подходящие методы решения систем уравнений с вырожденными матрицами.

3. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

Рассмотренный подход был реализован в среде графического программирования *LabVIEW* [10] и успешно применен для исследования влияния техники пилотирования и климатических условий базирования на интенсивность накопления повреждений в летательных аппаратах [1].

Соответствующая факторная модель показана на рис. 11. Для анализа интенсивности накопления повреждений и результатов психологического тестирования использовались только четыре последних по порядку из восьми вейвлет-коэффициентов, при анализе социально-экономических данных – последние два вейвлет-коэффициента из четырех, что было обусловлено особенностями прикладных задач. В модели, показанной на рис. 11,

представлены воздействия национальной техники пилотирования (факторы **R** и **F**) и воздействия условий эксплуатации (факторы **D** и **A**). Проведённый анализ выявил причинные и временные взаимосвязи между латентными факторами, а также уровень статистической значимости различных компонентов модели [1].



*Рис. 11. Модель для исследования влияния техники пилотирования и климатических условий базирования на интенсивность накопления повреждений в летательных аппаратах: воздействия национальной техники пилотирования представлены факторами **R** и **F**, воздействия условий эксплуатации – факторами **D** и **A**. V_* – дисперсии, C_{**} и C_{***} – ковариации между факторами.*

4. ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Предложен новый метод выявления и исследования факторов, определяющих развитие сложных систем, который опирается на возможности вейвлет-преобразований и идентифицируемых факторных структур. Согласно предложенному подходу, значения коэффициентов, полученные в результате дискретного вейвлет-преобразования временного ряда наблюдаемого процесса и соответствующие различным периодам наблюдений, рассматриваются как значения наблюдаемых переменных в последующем конфирматорном факторном анализе, который, в свою очередь, используется для выявления динамики факторных влияний и оценки показателей взаимодействий между факторами.
2. Идентификация свободных параметров факторной модели (как правило, факторных дисперсий и ковариаций) выполняется с помощью новой прямой (неитерационной) процедуры, опирающейся на метод максимального правдоподобия, что является альтернативой традиционному итерационному поиску локального решения задачи многомерной численной оптимизации, результат которого зависит от начального приближения. Основными особенностями этой процедуры являются:
 - составление переопределенной системы алгебраических уравнений относительно свободных параметров факторной модели с её последующим решением методом максимального правдоподобия;

- использование путевой модели дисперсионных составляющих;
 - использование вместо традиционного уровня значимости нового показателя для проверки гипотезы о представимости наблюдаемых результатов исследуемой факторной моделью, представляющего собой критический процент среднеквадратических отклонений компонентов вектора невязок от соответствующих компонентов наблюдаемых дисперсий и ковариаций;
 - возможность построения заключений о статистической значимости различных компонентов модели, используя статистические критерии согласия.
3. Сравнение различных видов факторных структур выявило преимущества путевой модели дисперсионных составляющих. Этот факт обусловлен линейностью аналитического представления дисперсий и ковариаций наблюдаемых переменных, что удобно для прямых оценок свободных параметров, а также большей степенью переопределенности модели.
 4. Для устранения возможной сингулярности исследуемой модели, приводящей к неоднозначности решения, разработан специальный метод выявления зависимых свободных параметров, сводящийся к решению алгебраической проблемы собственных значений и вращению базиса в невырожденном собственном подпространстве.
 5. Предложенный подход имеет существенные преимущества перед симплекс-методом, который обычно используется для исследования факторов, определяющих эволюцию системы.
 6. Предложенный подход реализован в среде графического программирования и применен при решении ряда прикладных задач.

ЛИТЕРАТУРА

1. Baranov S.N., Kuravsky L.S., Baranov N.I. *Studying influence of maneuvering loads occurrences and climatic conditions of basing on aircraft damage accumulation rate with the aid of trained structures*. In: Proc. 5th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Edinburgh, United Kingdom, July 2008, pp.118-126.
2. Bishop Y.M.M., Fienberg S.E., Holland P.W. *Discrete multivariate analysis: Theory and practice*. (Cambridge, MA, M I T Press, 1975.)
3. Bollen K.A. *Structural equations with latent variables*. (New York, John Wiley, 1989.)
4. Goldstein H. *Multilevel statistical models*. (3rd ed., London, Arnold, 2003.)
5. Jöreskog K.G. *Estimation and testing of simplex models*. British Journal of Mathematical and Statistical Psychology, Vol. 23, pp.121-145, 1970.
6. Kuravsky L.S. and Baranov S.N. *Development of the wavelet-based confirmatory factor analysis for monitoring of system factors*. – In: Proc. 5th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Edinburgh, United Kingdom, July 2008, pp.818-834.
7. Kuravsky L.S., Kornienko P.A. *On the approach to identifying quantitative trait loci in behavior genetics*. In: Proc. 2nd World Congress on Engineering Asset Management and 4th International Conference on Condition Monitoring, Harrogate, United Kingdom, pp.1133-1146, June 2007.
8. Куравский Л.С., Корниенко П.А. *Применение нейронных сетей для идентификации локусов количественных признаков в психогенетике*. - Нейрокомпьютеры: разработка и применение, №4, 2007, с. 15-31.
9. *LabVIEW tutorial for Windows*. (National Instruments Corp., 1994-2007.)
10. Лоули Д., Максвелл А. *Факторный анализ как статистический метод*. – М. Мир, 1967.
11. Loehlin J.C. *Latent variable models: An introduction to factor, path, and structural analysis*. (Hillsdale, NJ, Erlbaum, 1987.)
12. Neale M.C., Cardon L.R. *Methodology for genetic studies of twins and families*. (Dordrecht, the Netherlands, Kluwer Academic Publishers, 1992.)

13. Корольюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. *Справочник по теории вероятностей и математической статистике*. – М.: Наука, 1985.