

УДК 530.182

МОДЕЛИРОВАНИЕ СЛОЖНЫХ СИСТЕМ ПО НАБЛЮДАЕМЫМ ДАННЫМ НА ОСНОВЕ РЕКУРРЕНТНЫХ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

А. Ф. Селезнев^{1*}, А. С. Гаврилов¹, Д. Н. Мухин¹, Е. М. Лоскутов¹, А. М. Фейгин^{1,2}

¹ Институт прикладной физики РАН;

² Нижегородский госуниверситет им. Н. И. Лобачевского, г. Нижний Новгород, Россия

Предложен новый подход к реконструкции сложных пространственно-распределённых систем на основе порождаемых ими временных рядов. Он позволяет объединить два основных шага такой реконструкции — выбор набора фазовых переменных системы по наблюдаемым временным рядам и построение оператора эволюции, действующего в выбранном фазовом пространстве, с помощью искусственной нейронной сети со специальной топологией. Данная сеть, во-первых, отображает исходные высокоразмерные данные в пространство меньшей размерности и, во-вторых, задаёт оператор эволюции в этом пространстве. Эффективность подхода продемонстрирована на примере реконструкции системы Лоренца — высокоразмерной модели атмосферной динамики.

ВВЕДЕНИЕ

Построение динамических моделей путём прямого анализа данных (эмпирическое моделирование) наиболее актуально в случае сложных систем, уравнения для описания которых отсутствуют или не могут быть эффективно применены из-за больших неопределённостей в их параметрах; такая ситуация характерна для большинства живых, климатических, социо-экономических систем и т. д. Продолжающийся в различных областях наук прогресс в технологиях мониторинга и обработки данных позволяет проводить измерения с высоким разрешением (например, данные визуализации активности популяций нейронов или спутниковые измерения различных климатических характеристик), которые достаточны для реконструкции важнейших свойств динамической системы, лежащей в их основе. При этом естественные ограничения, связанные с дискретностью и конечностью анализируемых временных рядов, определяют диапазон временных масштабов, доступный для моделирования. Математически это означает, что реконструкция динамической системы возможна лишь в некотором подпространстве с размерностью, меньшей, чем размерность полного фазового пространства системы. Таким образом, эмпирические модели строятся в низкоразмерных проекциях полного фазового пространства сложной системы и призваны описывать её главные моды, дающие основной вклад в наблюдаемую динамику на рассматриваемых временных масштабах.

К настоящему времени проведено много исследований по разработке и приложениям различных подходов к реконструкции уравнений динамики системы по наблюдаемым временным рядам (см., например, книгу [1], а также работы [2–6]). Были созданы и успешно применены методы, эффективные для систем, принадлежащих различным частным классам: систем с известным оператором эволюции и скрытыми переменными [7–10], неавтономных систем [11–14], систем с запаздыванием [15], ансамблей связанных систем, в том числе с запаздыванием (см., например, [16, 17]), с разреженными матрицами связей [18–20], нейроосцилляторов [21, 22] и др.

Данная работа фокусируется на ситуации, характерной, в частности, для геофизических систем: высокая размерность входных данных, отсутствие информации как о размерности оператора эволюции (вообще говоря, она может быть бесконечной), так и об уравнениях динамики,

* aseleznev@ipfran.ru

наличие медленных изменений (трендов) параметров системы с масштабами, превышающими длину ряда, которые могут приводить к качественным изменениям наблюдаемого поведения — сменам динамических режимов. Например, такая ситуация возникает при эмпирическом моделировании климатической динамики: пространственно-распределённые данные отражают много-масштабный процесс, протекающий на фоне глобальных трендов (температуры, содержания углекислого газа и др.). Фактически, целью данной работы является попытка обобщения методов моделирования и прогноза качественного поведения слабонеавтономных динамических систем, разработанных авторами ранее [13, 23], на случай анализа высокоразмерных данных.

Описанная ситуация приводит к тому, что первый необходимый шаг построения эмпирической модели — реконструкция фазового пространства для построения оператора эволюции — становится нетривиальным. Например, в случае анализа климатических данных размерность временного ряда равна произведению числа узлов пространственной сетки на Земном шаре, в которых проводятся измерения, и числа измеряемых характеристик; обычно эта величина достигает десятков и сотен тысяч. При этом максимальная размерность восстанавливаемого по данным фазового пространства, в котором можно получить статистически значимую аппроксимацию оператора эволюции, сильно ограничена из-за конечной длины временного ряда. В этой связи чрезвычайно важной является задача предварительного уменьшения размерности наблюдаемых данных с целью выделения относительно небольшого числа переменных, содержащих информацию о ключевых свойствах наблюдаемой динамики.

Для построения таких переменных в настоящее время широко используются методы, основанные на различных линейных преобразованиях данных в пространственном домене. Наиболее известный из них — метод главных компонент [24] — позволяет представить наблюдаемый вектор данных в виде линейной суперпозиции ортогональных векторов. Получаемые при этом коэффициенты разложения — главные компоненты — являются некоррелирующими друг с другом временными рядами и могут быть использованы в качестве новых переменных. Фактически при такой процедуре исходный высокоразмерный временной ряд линейно отображается (проецируется) в пространство меньшей размерности. Существуют и более сложные линейные преобразования: поворот «варимакс» [25] — ортогональный поворот, применяемый для выявления максимально разделённых в пространстве компонент; метод независимых подпространств [26, 27], в котором требование некоррелированности главных компонент заменено требованием их независимости в смысле взаимной информации, и др. Кроме того, применяются различные пространственно-временные расширения данных методов, например мультисканальный сингулярный спектральный анализ [28]. Существенным недостатком описанных разложений является то, что они, в силу своей линейности, не могут адекватно учесть наличие нелинейных зависимостей в исследуемой системе и, следовательно, оказываются недостаточно эффективными для уменьшения размерности при наличии сильных нелинейных связей между временными рядами в различных пространственных точках. В этой связи в последнее время активно развиваются различные методы нелинейной редукции данных.

Некоторые из этих методов [29, 30] основаны на поиске новых переменных посредством применения стандартного метода главных компонент в пространстве с нелинейной метрикой. Их основным недостатком часто является отсутствие обратного преобразования из пространства новых переменных в пространство данных, которое является определяющим в практических приложениях. Другая группа методов основана на различных нелинейных обобщениях метода главных компонент с помощью искусственных нейронных сетей (ИНС). Так, автоассоциативные ИНС [31, 32] позволяют одновременно получить как низкоразмерную нелинейную проекцию данных, так и её обратное отображение в исходное пространство. А ИНС с обучающимися входами [33] задают лишь обратное преобразование, в то время как сами новые переменные получаются

путём оптимизации сигнала на входе. В этом направлении авторами данной работы был разработан метод, позволяющий находить главные нелинейные динамические моды, лежащие в основе эволюции пространственно-распределённой системы [34, 35].

Вторым важнейшим шагом является построение модели оператора эволюции, действующего в восстановленном фазовом пространстве низкоразмерной подсистемы. Основные сложности этого шага связаны с тем, что в общем случае конкретный вид оператора эволюции исследуемой системы неизвестен и его приходится искать в некотором широком классе функций, обладающих свойствами универсального аппроксиматора. Данной проблеме посвящено огромное число работ (см., например, работу [36] и обзоры [37, 38]), в которых для параметризации оператора эволюции используются различные классы аппроксиматоров, в частности полиномиальные функции и ИНС [13, 14, 39–42].

Таким образом, успех в решении задачи построения эмпирической модели сложной системы определяется, во-первых, правильным выбором фазовых переменных, получаемых в процессе редукции исходных высокоразмерных данных, и, во-вторых, оптимальной параметризацией оператора эволюции, действующего в этом фазовом пространстве. Как правило, насколько известно авторам, эти шаги делаются либо независимо друг от друга, либо совместно, но в ограниченном (например, линейном [43]) варианте, что существенно сужает возможности такого описания.

В данной работе предлагается рекуррентная ИНС со специальной топологией, позволяющая совместить нелинейную редукцию размерности и построение нелинейного оператора эволюции. Данная ИНС разработана на основе автоассоциативной ИНС, предложенной в работе [31]. Такая сеть состоит из трёх внутренних слоёв и отображает пространство исходных данных в себя. Первый слой отображает входной вектор данных высокой размерности в пространство меньшей размерности — пространство скрытых состояний системы; во втором слое формируется вектор уменьшенной размерности. Данный переход называется «mapping» [31], он позволяет получить низкоразмерную проекцию исходного пространства. В третьем слое происходит обратное отображение в пространство наблюдаемых данных (шаг «demapping»). В данной работе мы сделали второй слой сети динамическим, т.е., во-первых, разместили на нём нейроны с нелинейной функцией активации и, во-вторых, выход этого слоя связали не только со следующим слоем (demapping), но и с собственным входом. Таким образом, мы получили рекуррентную ИНС, обладающую собственной динамикой и способную генерировать временные ряды. Обучать такую сеть можно, «предъявляя» ей различные начальные состояния (последовательности векторов данных) на входе и следующие по времени серии векторов данных на выходе.

Для обучения сети используется байесов формализм решения обратных задач: строится ценовая функция для параметров ИНС в форме апостериорной плотности вероятности, максимум которой соответствует обученной ИНС. Используемая ценовая функция, основанная на идеях работ [10, 44], направлена на получение модели, фазовые траектории которой «держатся» в окрестности фазовой траектории системы на сегментах максимально возможной протяжённости (в случае хаотической динамики протяжённость сегментов ограничена величиной обратного старшего ляпуновского показателя системы).

Стоит отметить, что рекуррентные ИНС становятся всё более популярны в задачах реконструкции сложных систем по временным рядам. Используются различные топологии сетей и методы их обучения, основанные на стандартных алгоритмах обратного распространения ошибки [45, 46], расширенном фильтре Калмана [47], активно развивающихся алгоритмах резервуарных вычислений [48–50] и других алгоритмах [51, 52]. Топология ИНС, предлагаемая в данной работе, предназначена для эффективного моделирования динамической системы в ситуации, когда есть основания предполагать наличие в измерениях сигнала от низкоразмерной подсистемы, который «зашифрован» в высокоразмерных данных посредством некоторой неизвестной функции.

В разделе 1 данной статьи мы формализуем конкретный вид модели в форме многослойной рекуррентной ИНС. В разделе 2 описывается способ обучения предложенной ИНС в рамках байесова подхода. В разделе 3 приведены основные сведения о системе Лоренца, используемой здесь в качестве источника данных для эмпирической реконструкции при иллюстрации возможностей ИНС. В разделе 4 демонстрируется способность предложенной ИНС к моделированию и прогнозу динамических режимов в системе Лоренца с медленно меняющимся управляющим параметром. В разделе 5 обсуждаются полученные результаты и перспективы дальнейшего развития предложенной методики.

1. ЭМПИРИЧЕСКАЯ МОДЕЛЬ В ФОРМЕ МНОГОСЛОЙНОЙ РЕКУРРЕНТНОЙ ИНС

Пусть имеется временной ряд $\mathbf{x}_0, \dots, \mathbf{x}_{N-1}$, где $\mathbf{x}_n \in \mathbb{R}^D$, представляющий измерения в эквидистантные моменты времени $t_0, \dots, t_{N-1}$ в D точках на пространственной сетке. Тогда, согласно парадигме, сформулированной во Введении, будем строить эмпирическую модель, сочетающую в себе три основных шага:

- 1) отображение исходных данных в пространство меньшей размерности d : $\mathbf{p}_n = \varphi(\mathbf{x}_n)$; $\varphi : \mathbb{R}^D \rightarrow \mathbb{R}^d$, $\mathbf{p}_n \in \mathbb{R}^d$, $d < D$ (mapping);
- 2) оператор эволюции (вообще говоря, неавтономный), действующий в редуцированном пространстве новых переменных $\mathbf{p}$: $\mathbf{p}_{n+1} = \mathbf{Q}(\mathbf{p}_n, t_n)$, $\mathbf{p}_{n+2} = \mathbf{Q}(\mathbf{p}_{n+1}, t_{n+1}), \dots$; $\mathbf{Q}(\cdot, t) : \mathbb{R}^d \rightarrow \mathbb{R}^d$;
- 3) отображение в исходное пространство данных: $\mathbf{x}_n = \psi(\mathbf{p}_n)$; $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^D$ (demapping).

Функции $\varphi, \mathbf{Q}, \psi$ параметризуются с помощью трёхслойных персептронов, которые являются универсальными аппроксиматорами произвольных нелинейных зависимостей [53]. Структура персептрона устроена следующим образом: сигнал со входного слоя подаётся на скрытый слой, который представлен нейронами с нелинейной функцией активации (в данной работе используется гиперболический тангенс), затем на выходном слое формируется линейная комбинация этих нейронов. Для функций $\varphi, \mathbf{Q}, \psi$ можно записать следующие явные выражения:

$$\begin{aligned}\varphi(\mathbf{x}_n) &= \sum_{i=1}^{m_\varphi} \alpha_i^\varphi \tanh [(\omega_i^\varphi)^T \mathbf{x}_n + \gamma_i^\varphi], \\ \mathbf{Q}(\mathbf{p}_n, t_n) &= \sum_{i=1}^{m_Q} \left[\alpha_i^Q + \beta_i^Q \left(\frac{t_n}{N-1} \right) \right] \tanh [(\omega_i^Q)^T \mathbf{p}_n + \gamma_i^Q], \\ \psi(\mathbf{p}_n) &= \sum_{i=1}^{m_\psi} \alpha_i^\psi \tanh [(\omega_i^\psi)^T \mathbf{p}_n + \gamma_i^\psi],\end{aligned}\tag{1}$$

где $\alpha_i^\varphi \in \mathbb{R}^d$, $\alpha_i^Q \in \mathbb{R}^d$, $\alpha_i^\psi \in \mathbb{R}^D$, $\beta_i^Q \in \mathbb{R}^d$, $\omega_i^\varphi \in \mathbb{R}^D$, $\omega_i^Q \in \mathbb{R}^d$, $\omega_i^\psi \in \mathbb{R}^d$, $\gamma_i^\varphi \in \mathbb{R}$, $\gamma_i^Q \in \mathbb{R}$, $\gamma_i^\psi \in \mathbb{R}$ — параметры, m_φ, m_Q, m_ψ — числа нейронов в скрытых слоях персептронов.

Выбор количества нейронов в разных слоях определяет допустимую степень нелинейности соответствующих блоков. Так, m_φ задаёт возможную нелинейность проекции данных на низкоразмерное многообразие модели, определяя сложность низкоразмерного сигнала в последующем, рекуррентном, слое ИНС. Параметр m_Q отвечает за сложность редуцированного оператора эволюции системы, определяющего динамические свойства модели. Параметр последнего слоя m_ψ задаёт допустимую сложность преобразования из низкоразмерного многообразия модели в пространство данных, определяя тем самым нелинейность вложения этого многообразия в пространство данных и напрямую влияя на точность аппроксимации данных низкоразмерным многообра-

зием (чрезмерно большое m_ψ может привести к точной аппроксимации любого, даже случайного, ряда за счёт «изрезанности» отображения).

В выходной слой персептрона, параметризующего оператор эволюции $\mathbf{Q}$, введена явная зависимость от времени. Здесь t_n — дискретное время (фактически номер момента времени), причём $t_n/(N-1)$ изменяется в диапазоне от 0 до 1 на длине наблюдаемого временного ряда. Как было показано в работах [13, 14, 41, 42], такая аппроксимация оператора эволюции является эффективной при реконструкции слабонеавтономных динамических систем. В конечной форме модель представляет собой многослойную рекуррентную ИНС:

$$\mathbf{x}_{n+j} = \mathbf{F}^{(j)}(\mathbf{x}_n, \boldsymbol{\mu}, t_n) + \mathbf{z}_n. \quad (2)$$

Здесь $\boldsymbol{\mu}$ — вектор, получаемый постановкой в один столбец всех параметров персептронов, соответствующих функциям $\varphi, \mathbf{Q}, \psi$, причём его размерность равна полному числу этих параметров; $\mathbf{F}^{(j)}(\mathbf{x}_n, \boldsymbol{\mu}, t_n) = \psi\{\mathbf{Q}^{(j)}[\varphi(\mathbf{x}_n), t_n]\}$, где j — число применений оператора эволюции в редуцированном пространстве новых переменных, т. е. $\mathbf{Q}^{(0)}[\varphi(\mathbf{x}_n), t_n] = \varphi(\mathbf{x}_n)$, $\mathbf{Q}^{(j)}[\varphi(\mathbf{x}_n), t_n] = \mathbf{Q}\{\mathbf{Q}^{(j-1)}[\varphi(\mathbf{x}_n), t_n], t_{n+j}\}$; $\mathbf{z}_n \in \mathbb{R}^D$ — погрешность аппроксимации, которая, по аналогии с работами [34, 35], предполагается гауссовой, т. е. $\mathbf{z}_n = \sigma \boldsymbol{\xi}_n$, $\boldsymbol{\xi}_n \in \mathbb{R}^D$ — вектор, компоненты которого представляют собой независимые дельта-коррелированные гауссовы случайные процессы с единичной дисперсией.

2. ОБУЧЕНИЕ РЕКУРРЕНТНОЙ ИНС В РАМКАХ БАЙЕСОВА ПОДХОДА

После того, как определена явная форма эмпирической модели (2), задача реконструкции сводится к нахождению набора её неизвестных параметров $\boldsymbol{\mu}$ и σ . Ясно, что, поскольку ИНС обладают свойствами универсальных аппроксиматоров, существует бесконечно много моделей различной сложности в форме (2), способных воспроизвести имеющиеся данные с любой наперёд заданной точностью. Другими словами, нахождение неизвестных параметров модели $\boldsymbol{\mu}, \sigma$ при фиксированных m_φ, m_Q, m_ψ и d является некорректной обратной задачей. В данной работе они определяются в рамках байесова подхода, эффективность которого при решении задач обратного моделирования была продемонстрирована в работах [13, 14, 34, 35, 41]. В качестве ценовой функции используется апостериорная плотность вероятности неизвестных параметров модели, которая, согласно теореме Байеса, пропорциональна произведению их априорных плотностей вероятности и условной плотности вероятности наблюдаемых данных (правдоподобия):

$$P(\boldsymbol{\mu}, \sigma | \mathbf{x}_0, \dots, \mathbf{x}_{N-1}) \propto P_{\text{pr}}(\boldsymbol{\mu}) P_{\text{pr}}(\sigma) P(\mathbf{x}_0, \dots, \mathbf{x}_{N-1} | \boldsymbol{\mu}, \sigma). \quad (3)$$

При этом оптимальные значения параметров $\boldsymbol{\mu}, \sigma$ соответствуют максимуму функции (3). На практике задача максимизации функции (3) сводится к минимизации (по $\boldsymbol{\mu}, \sigma$) функции

$$L(\boldsymbol{\mu}, \sigma) = -\ln [P(\boldsymbol{\mu}, \sigma | \mathbf{x}_0, \dots, \mathbf{x}_{N-1})]. \quad (4)$$

С учётом того, что погрешности $\mathbf{z}_n$ в (2) являются гауссовыми и дельта-коррелированными, $P(\mathbf{x}_0, \dots, \mathbf{x}_{N-1} | \boldsymbol{\mu}, \sigma)$ в правой части (3) даёт формулой

$$P(\mathbf{x}_0, \dots, \mathbf{x}_{N-1} | \boldsymbol{\mu}, \sigma) = \prod_{j=0}^{N-1} \frac{1}{(2\pi\sigma^2)^{D/2}} \exp \left[-\frac{\|\mathbf{x}_j - \mathbf{F}^{(j)}(\mathbf{x}_0, \boldsymbol{\mu}, t_0)\|^2}{2\sigma^2} \right]. \quad (5)$$

Как отмечалось в работе [10], в случае достаточно протяжённого временного ряда зависимость правдоподобия (5) от своих аргументов имеет сложную мультимодальную форму, что сильно затрудняет оптимизацию ценовой функции (3). В работе [10] был предложен способ модификации

правдоподобия (5), основанный на предположении, что состояния системы в достаточно разнесённые друг от друга моменты времени статистически независимы, а динамическая связь между ними сохраняется лишь внутри сегментов с протяжённостью w . Такая ситуация является типичной для хаотических систем, которые «забывают» своё начальное состояние через промежуток времени, обратно пропорциональный старшему ляпуновскому показателю. С учётом сказанного, выражение (5) факторизуется следующим образом:

$$P(\mathbf{x}_0, \dots, \mathbf{x}_{N-1} | \boldsymbol{\mu}, \sigma) = \prod_{n=T+w}^{N-1} \prod_{k=1}^D \frac{1}{\sqrt{2\pi\lambda_k}} \exp\left(-\frac{x_{k,n}^2}{2\lambda_k}\right) \prod_{n=0,w,\dots,T} \prod_{j=0}^{w-1} \frac{1}{(2\pi\sigma^2)^{D/2}} \exp\left[-\frac{\|\mathbf{x}_{n+j} - \mathbf{F}^{(j)}(\mathbf{x}_n, \boldsymbol{\mu}, t_n)\|^2}{2\sigma^2}\right], \quad (6)$$

где $\lambda_k = N^{-1} \sum_{n=0}^{N-1} x_{k,n}^2$ (без ограничения общности здесь и далее мы полагаем, что временной ряд центрирован: $N^{-1} \sum_{n=0}^{N-1} x_{k,n} = 0$). Таким образом, при обучении ИНС оператор эволюции действует в пределах непересекающихся сегментов с длиной w , на которые делится исходный ряд наблюдаемых данных с длиной N , причём $\mathbf{x}_n$ — начальные точки сегментов, $n = 0, w, 2w, \dots, T$, $T \leq N - w + 1$. Два множителя в выражении (6) отражают тот факт, что в общем случае на интервале длины N укладывается $[N/w] = T/w$ сегментов с длиной w , на которых действует модель, а оставшиеся $N - T - w + 1 < w$ отсчётов ряда трактуются как параметры модели (подробнее об этом см. в работе [54]). Отметим, что при $w = 1$ модель (2) переходит в стандартную автоассоциативную ИНС, описанную в работе [31].

По аналогии с работами [13, 14, 34, 35, 41] в качестве плотности вероятности $P_{\text{pr}}(\boldsymbol{\mu})$ мы используем гауссовы распределения с дисперсиями $\sigma_{\alpha\varphi}^2, \sigma_{\alpha Q}^2, \dots$ для каждого параметра из вектора $\boldsymbol{\mu}$ (при этом все параметры ИНС полагаются априори взаимно независимыми; таким образом, величина $P_{\text{pr}}(\boldsymbol{\mu})$ является произведением гауссовых плотностей вероятности каждого параметра). Значения $\sigma_{\alpha\varphi}, \sigma_{\alpha Q}, \dots$ фактически задают область изменения соответствующих параметров в процессе обучения ИНС. В этой работе они выбираются следующим образом: $\sigma_{\alpha\varphi} = 1/\sqrt{m_\varphi}$, $\sigma_{\alpha Q} = 1/\sqrt{m_Q}$, $\sigma_{\beta Q} = 1/\sqrt{m_Q}$, $\sigma_{\alpha\psi,k} = \sqrt{\lambda_k/m_\psi}$, $\sigma_{\omega\varphi} = 1/\sqrt{\sum_{k=1}^D \lambda_k}$, $\sigma_{\omega Q} = 1/\sqrt{d}$, $\sigma_{\omega\psi} = 1/\sqrt{d}$, $\sigma_{\gamma\varphi} = \sigma_{\omega\varphi} \sqrt{D}$, $\sigma_{\gamma Q} = \sigma_{\omega Q} \sqrt{d}$, $\sigma_{\gamma\psi} = \sigma_{\omega\psi} \sqrt{d}$. В основе такого выбора лежат соображения о равенстве вариаций входов и выходов в каждом слое ИНС, обобщённые на случай некоррелирующих компонент ряда $\mathbf{x}$ с произвольной вариацией. Как было показано в работе [55], при данных значениях $\sigma_{\alpha\varphi}, \sigma_{\alpha Q}, \dots$ удаётся частично избежать проблемы «затухающих» и «взрывающихся» градиентов [56], возникающей в процессе обучения многослойных ИНС с использованием градиентных алгоритмов оптимизации. В качестве априорной плотности вероятности параметра σ , как и в работах [13, 34, 35], используется неинформативное распределение $P_{\text{pr}}(\sigma) = \text{const}$.

Для минимизации ценовой функции (4) в данной статье использовалась комбинация двух известных численных методов — алгоритма стохастического градиентного спуска (в модификации Adam [57]) и квазиньютоновского метода Бройдена—Флетчера—Гольдфарба—Шанно (БФГШ) [58]. Предварительно ценовая функция масштабировалась таким образом, чтобы характерные значения, принимаемые ею, были порядка 1. При каждой фиксированной комбинации структурных параметров ИНС ($m_\varphi, m_Q, m_\psi, d$) проводилось 1 000 актов минимизации масштабированной ценовой функции при случайных начальных условиях. При этом определялся набор оптимальных параметров ИНС, соответствующих наименьшему по всем актам минимизации значению ценовой функции. В каждом случае на первом этапе использовался алгоритм Adam, а критерием остановки являлось требование того, чтобы норма вектора, содержащего значения градиентов ценовой функции по всем параметрам, не превосходила 10^{-3} (в случае недостижимости этого условия

алгоритм останавливается после выполнения 25 000 итераций). Таким образом значения параметров приводились в окрестность минимума ценовой функции. На втором этапе использовался алгоритм БФГШ, характерное количество итераций которого составляло примерно 2 000. Для каждого результата минимизации ценовой функции проверялось равенство нулю нормы вектора её градиентов по всем параметрам (с точностью до 10^{-6}), а также положительная определённость матрицы её вторых производных.

3. СИСТЕМА ЛОРЕНЦА

Система Лоренца (в англоязычной литературе Lorenz-96) была предложена Эдвардом Лоренцем [59] для изучения фундаментальных вопросов, касающихся предсказуемости динамики атмосферы и прогнозирования погоды. Она записывается следующим образом:

$$\frac{dX_i}{dt} = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F, \quad i = 1, \dots, D, \quad D > 3, \quad X_{i \pm D} = X_i. \quad (7)$$

Уравнения (7) отражают основные факторы атмосферной динамики: квадратичные слагаемые описывают адвекцию, линейные — диссипацию, а константа F — внешнее воздействие [59]. Переменные X_i можно интерпретировать как некоторые характеристики атмосферы, такие как давление или завихренность, в различных точках i по долготе. Таким образом, система описывает атмосферные волны, распространяющиеся по кругу с фиксированной широтой на Земном шаре. В настоящее время модель Лоренца применяется для тестирования различных методов ассимиляции данных [60, 61], параметризации подсеточных процессов в моделях глобальной циркуляции атмосферы и океана [62], в статистической физике [63] и в общих исследованиях пространственно-временного хаоса [64].

Динамику системы определяют два управляющих параметра: число точек пространственной сетки D (в данном случае задающих размерность наблюдаемого вектора состояния системы) и величина константы внешнего воздействия F . Нетрудно видеть, что точка $\mathbf{X} = (F, \dots, F)$ является состоянием равновесия при любых значениях параметров D и F . При $F < 8/9$ динамика системы тривиальна (вне зависимости от D): существует единственное устойчивое решение $X_1 = \dots = X_D = F$ [65]. При увеличении F в системе появляются периодические решения, а затем (при определённом D) возникает хаос [59].

4. ПРОГНОЗ ДИНАМИЧЕСКИХ ПЕРЕХОДОВ В СИСТЕМЕ ЛОРЕНЦА

Динамика системы Лоренца рассматривалась при фиксированном значении параметра $D = 22$; параметр F изменялся в пределах от 5 до 3. Согласно результатам работы [66], при $F = 5$ в системе присутствует хаотический аттрактор с небольшим значением фрактальной размерности, а при $F = 3$ — устойчивый предельный цикл (аттрактор с размерностью 1).

Размерность аттрактора системы является чрезвычайно информативной характеристикой, позволяющей детектировать бифуркации в системе, связанные с изменением типа аттрактора. Резкое изменение размерности при медленном изменении управляющего параметра системы является индикатором такой бифуркации. В силу вышесказанного, мы будем использовать размерность аттрактора в качестве основной характеристики при сравнении качественных модельных предсказаний с поведением системы. Для расчёта размерности аттрактора мы используем определение корреляционной размерности множества (см., например, [67]), по причине простоты её численной оценки по имеющемуся временному ряду. В приложении к реальному временному

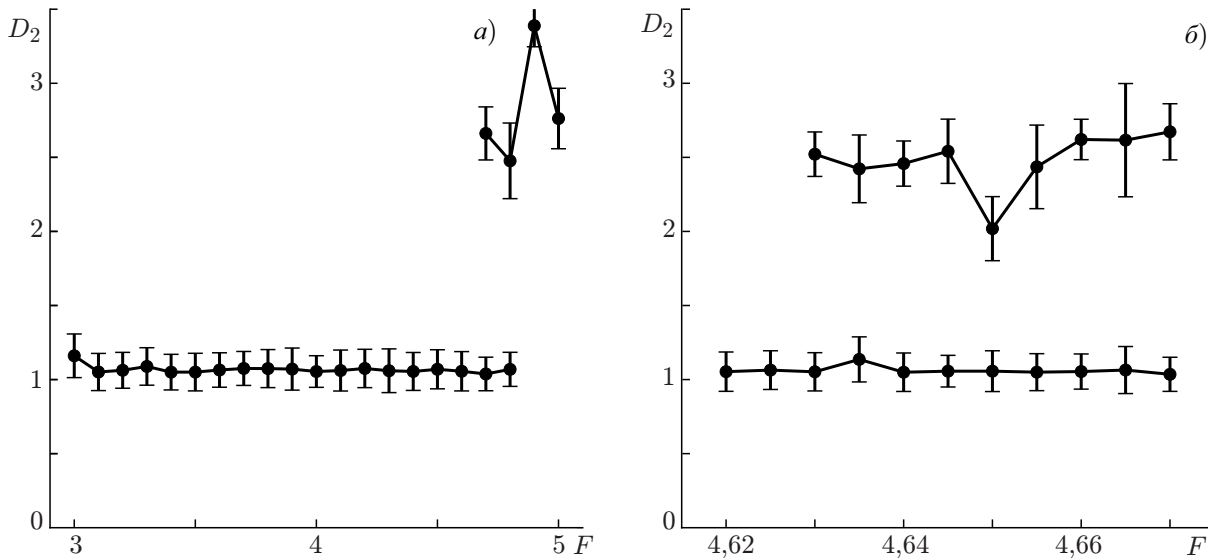


Рис. 1. Корреляционная размерность аттрактора системы Лоренца в зависимости от параметра F для $F = 3 \div 5$ с шагом 0,1 (а) и $F = 4,6 \div 4,7$ с шагом 0,01 (б). Доверительный интервал 65 %

ряду, имеющему ограниченную длину, корреляционной размерностью будем называть показатель D_2 в аппроксимации закона изменения оценки корреляционного интеграла C в зависимости от масштаба ϵ :

$$C(\epsilon) \propto \epsilon^{D_2}, \quad (8)$$

где $C(\epsilon)$ оценивается как доля пар векторов в фазовом пространстве, восстановленном по временному ряду, евклидово расстояние между которыми меньше ϵ в каждой паре. Подробное описание алгоритма оценки D_2 по временному ряду см. в работе [68].

На рис. 1 изображена корреляционная размерность D_2 аттрактора системы Лоренца при $D = 22$ в зависимости от параметра F . Для каждого значения F приведены средние значения корреляционной размерности, рассчитанные по 100 реализациям временного ряда, и их доверительные интервалы. В случае каждой реализации система интегрировалась при случайных начальных условиях в течение 100 000 временных отсчётов для достижения установившегося режима, а для расчёта корреляционной размерности использовались последние 10 000 отсчётов.

Как видно из рис. 1, при изменении параметра F от 5 до 3 среднее значение корреляционной размерности аттрактора системы уменьшается до 1, т. е. динамический режим меняется от сложного (хаос) к простому (предельный цикл); при этом в диапазоне F , приблизительно соответствующем изменению этого параметра от 4,8 до 4,6, в системе присутствует мультистабильность, о чём свидетельствует наличие двух «ветвей» в зависимости корреляционной размерности. Нами исследовалась возможность эмпирического прогноза данной смены режима с помощью ИНС, введённой в разделе 1. Для этого был сгенерирован 22-мерный нестационарный временной ряд с длиной в 10 000 отсчётов с шагом по времени 0,1 (см. рис. 2а). Параметр F при этом линейно уменьшался от 5 до 3. Эмпирическая модель (2) в форме рекуррентной ИНС строилась по начальному интервалу нестационарного временного ряда с длиной в 2 000 отсчётов, который соответствует изменению параметра F от 5 до 4,6 (отметим, что данный участок временного ряда отвечает нахождению системы в окрестности сложного режима). Обучающая выборка данных предварительно раскладывалась по базису эмпирических ортогональных функций (см. Введение). Таким образом, при обучении использовались некоррелирующие временные ряды, что важно с точки зрения выбранных априорных ограничений на параметры ИНС (см. раздел 2). Явная

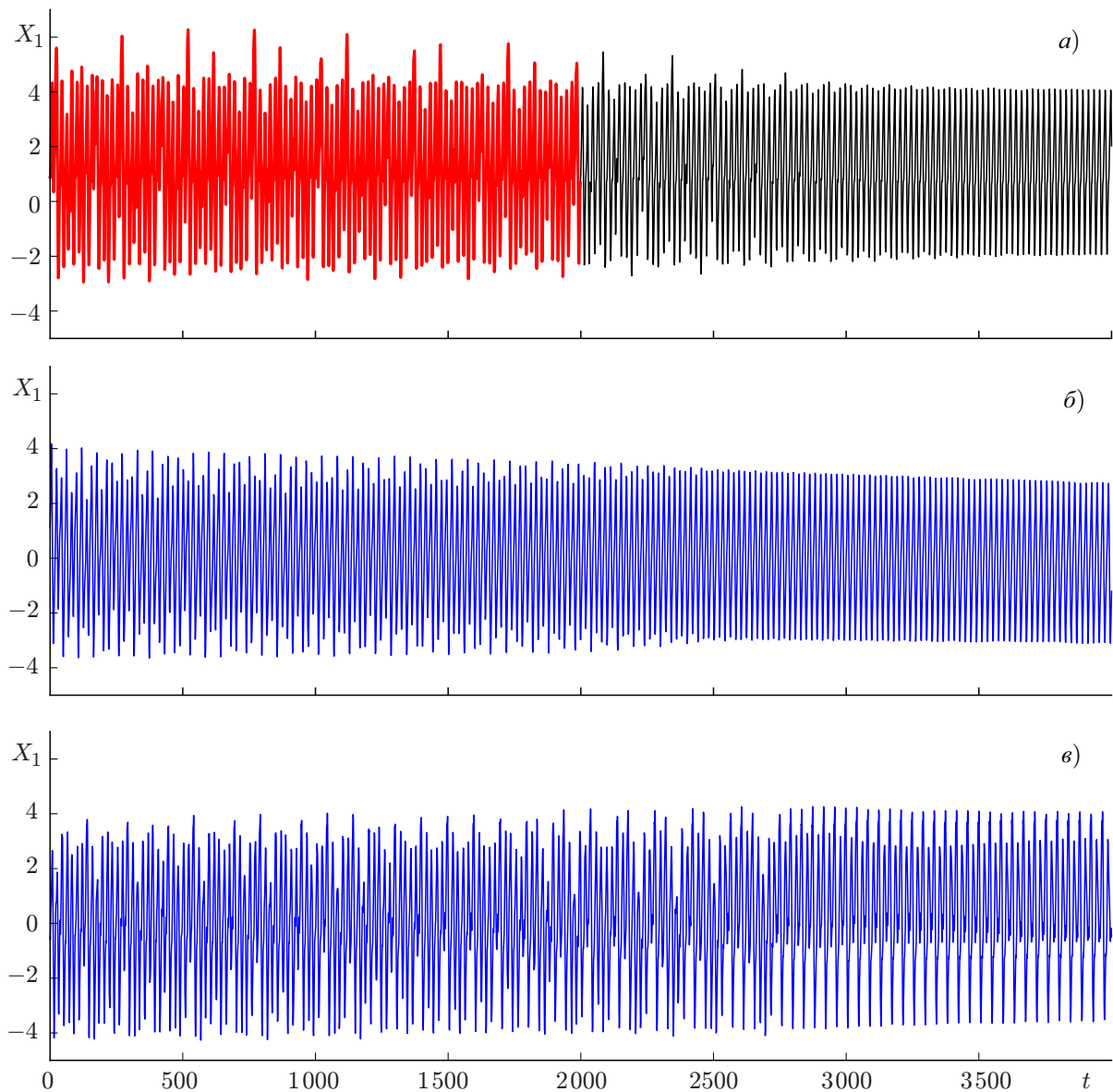


Рис. 2. Фрагмент нестационарного временного ряда для переменной X_1 системы Лоренца (а) и его воспроизведение ИНС (б и в). Значения структурных параметров ИНС: $m_\varphi = 8, m_Q = 6, m_\psi = 8, d = 5, w = 6$ (б) и $m_\varphi = 8, m_Q = 6, m_\psi = 28, d = 5, w = 6$ (в). Красным цветом на панели а отмечен участок временного ряда, по которому проводилось обучение ИНС

зависимость модели (2) от времени позволяет как проследить эволюцию динамического режима на интервале обучения, так и экстраполировать модель за пределы обучающей выборки и тем самым спрогнозировать качественное поведение системы.

Важным вопросом при обучении ИНС является оптимальный выбор её структуры, которая определяется числами нейронов во внутренних слоях m_φ, m_Q, m_ψ , размерностью редуцированного пространства d и величиной сегмента временного ряда w . Числа нейронов задают сложность нелинейных функций, которые могут быть аппроксимированы ИНС. Ясно, что при очень большом числе нейронов можно добиться сколь угодно точных аппроксимаций зависимостей $\mathbf{x}_{n+1}$ от $\mathbf{x}_n$ на интервале обучения. Однако такая модель будет «переобученной», т.е. не способной описывать поведение системы за пределами обучающей выборки. С другой стороны, при слишком

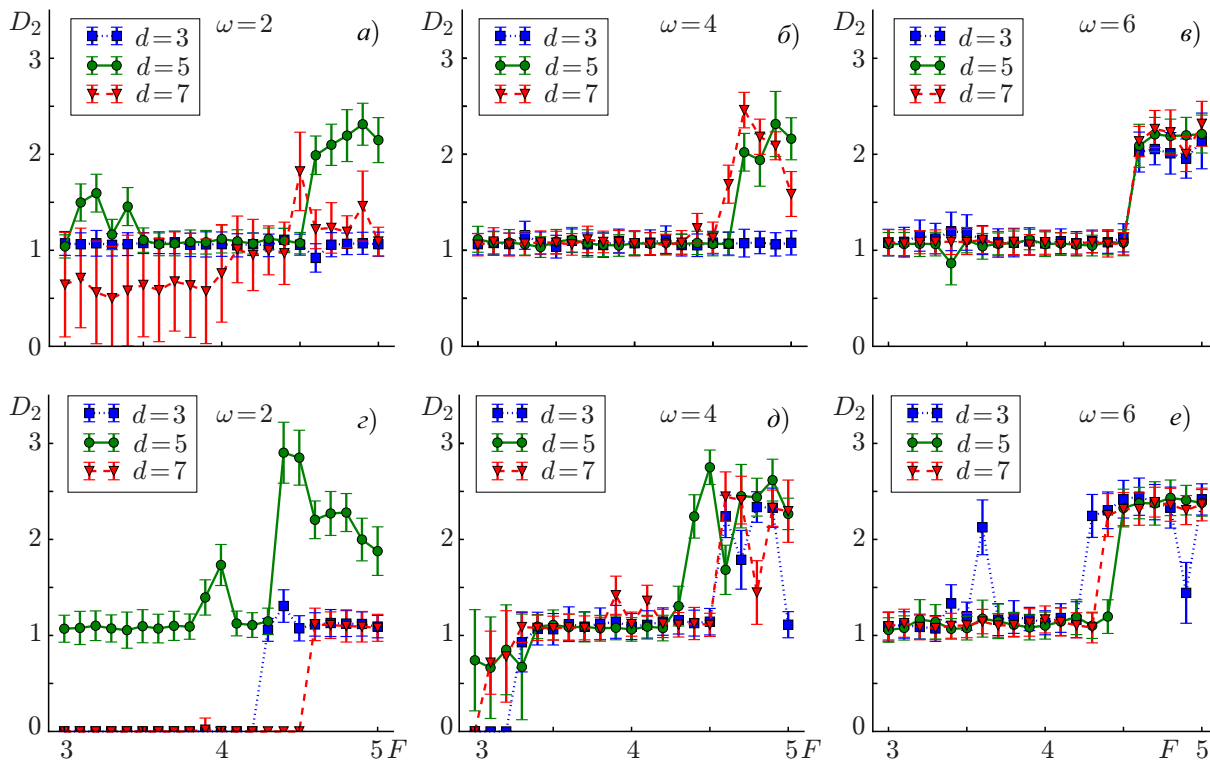


Рис. 3. Корреляционные размерности аттракторов ИНС при различных значениях её структурных параметров в зависимости от параметра F . Панели а–в: $m_\varphi = 8, m_Q = 6$ и $m_\psi = 8$, панели г–е: $m_\varphi = 8, m_Q = 6$ и $m_\psi = 28$. Доверительный интервал 65 %

малом количестве нейронов модель будет не в состоянии воспроизвести ключевые свойства наблюдаемой динамики, что необходимо для качественного прогноза. В данной работе в иллюстративных целях структурные параметры $m_\varphi = 8, m_Q = 6$ и $m_\psi = 8$ были выбраны из диапазона, соответствующего адекватному поведению модели. Кроме того, для сравнения показано поведение модели при параметрах $m_\varphi = 8, m_Q = 6$ и $m_\psi = 28$, отвечающего «переобученной» модели, для которой вложение низкоразмерного многообразия сильно нелинейно.

Размерность редуцированного фазового пространства d определяется размерностью вложения аттрактора моделируемой низкоразмерной подсистемы. Согласно теореме Мане [69], вложение аттрактора размерности M может быть обеспечено в пространстве размерности $d \geq 2\bar{M} + 1$, где $\bar{M}$ — ближайшее к M сверху целое число. Исходя из графика для корреляционной размерности (см. рис. 1), в качестве оценки размерности восстанавливаемого аттрактора системы Лоренца логично положить $\bar{M} = 3$. Тогда, согласно формуле Мане минимальная размерность вложения аттрактора $d = 2 \cdot 3 + 1 = 7$. Как отмечалось в работе [70], значение d , определяемое по формуле Мане, часто оказывается завышенным. Поэтому динамика ИНС исследовалась при значениях d в пределах от 3 до 7. Величина сегмента w является техническим параметром и определяет число применений оператора эволюции в редуцированном пространстве при обучении ИНС. Следовательно, её увеличение в некоторых пределах приводит к более устойчивому восстановлению фазовой траектории реконструируемой подсистемы. В данном случае w варьировалось от 2 до 6.

На рис. 3 изображены средние значения корреляционных размерностей аттракторов ИНС, соответствующих различным значениям управляющего параметра F системы Лоренца. Из графиков видно, что модель в форме ИНС, построенная по интервалу нестационарного 22-мерного временного ряда системы Лоренца, способна воспроизвести изменение корреляционной размерно-

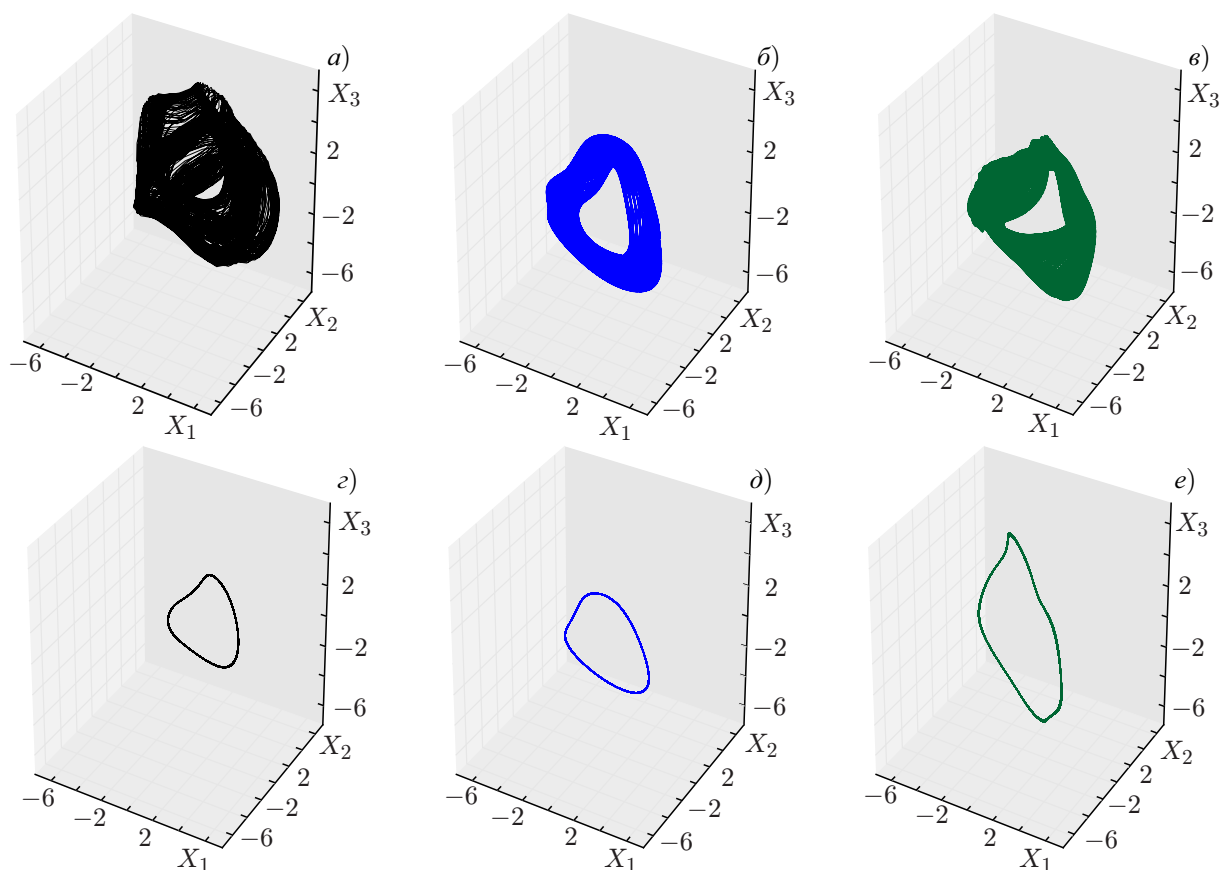


Рис. 4. Проекция аттракторов системы Лоренца при $F = 5$ (а), $F = 3$ (с) и проекция аттракторов ИНС, соответствующих $F = 5$ (б, е), $F = 3$ (д, е). Структурные параметры ИНС: $m_\varphi = 8$, $m_Q = 6$, $m_\psi = 8$, $d = 5$, $w = 6$ (б, д) и $m_\varphi = 8$, $m_Q = 6$, $m_\psi = 28$, $d = 5$, $w = 6$ (е, е)

сти её аттрактора в широком диапазоне значений управляющего параметра F . При этом качество воспроизведения существенно зависит от выбора структурных параметров ИНС. Зависимости на рис. 3 изображены при фиксированных числах нейронов в слоях ИНС $m_\varphi = 8$, $m_Q = 6$ и $m_\psi = 8$ (а–в), $m_\varphi = 8$, $m_Q = 6$ и $m_\psi = 28$ (г)–(е) и нескольких значениях величины сегмента $w = 2, 4, 6$ и размерности редуцированного пространства $d = 3, 5, 7$. Видно, что при минимальном значении параметра $w = 2$ (и любом из рассмотренных $d = 3, 5, 7$) ИНС не в состоянии описать зависимость корреляционной размерности аттрактора системы Лоренца от параметра F ни на интервале обучения (F от 5,0 до 4,6), ни за его пределами. При увеличении w точность воспроизведения заметно улучшается: ИНС качественно описывает зависимость корреляционной размерности аттрактора системы Лоренца как на интервале обучения, так и за его пределами, предсказывая её спад при $F \approx 4,6$. Выбор значения размерности редуцированного пространства d также играет важную роль. Так, при минимальном $d = 3$ ИНС некорректно описывает зависимость корреляционной размерности аттрактора системы Лоренца при всех $w = 2, 4, 6$, а наиболее удачное соответствие системы и модели достигается при $d = 5$; при дальнейшем увеличении d от 5 до 7 качество воспроизведения ИНС зависимости корреляционной размерности существенно не улучшается.

Таким образом, как следует из графиков на рис. 3, оптимальные значения структурных параметров (размерности d редуцированного фазового пространства и длины сегмента w временного ряда) ИНС, построенной по имеющемуся временному ряду системы Лоренца, равны $d = 5$ и $w = 6$ соответственно; следующие графики построены при данных значениях этих параметров.

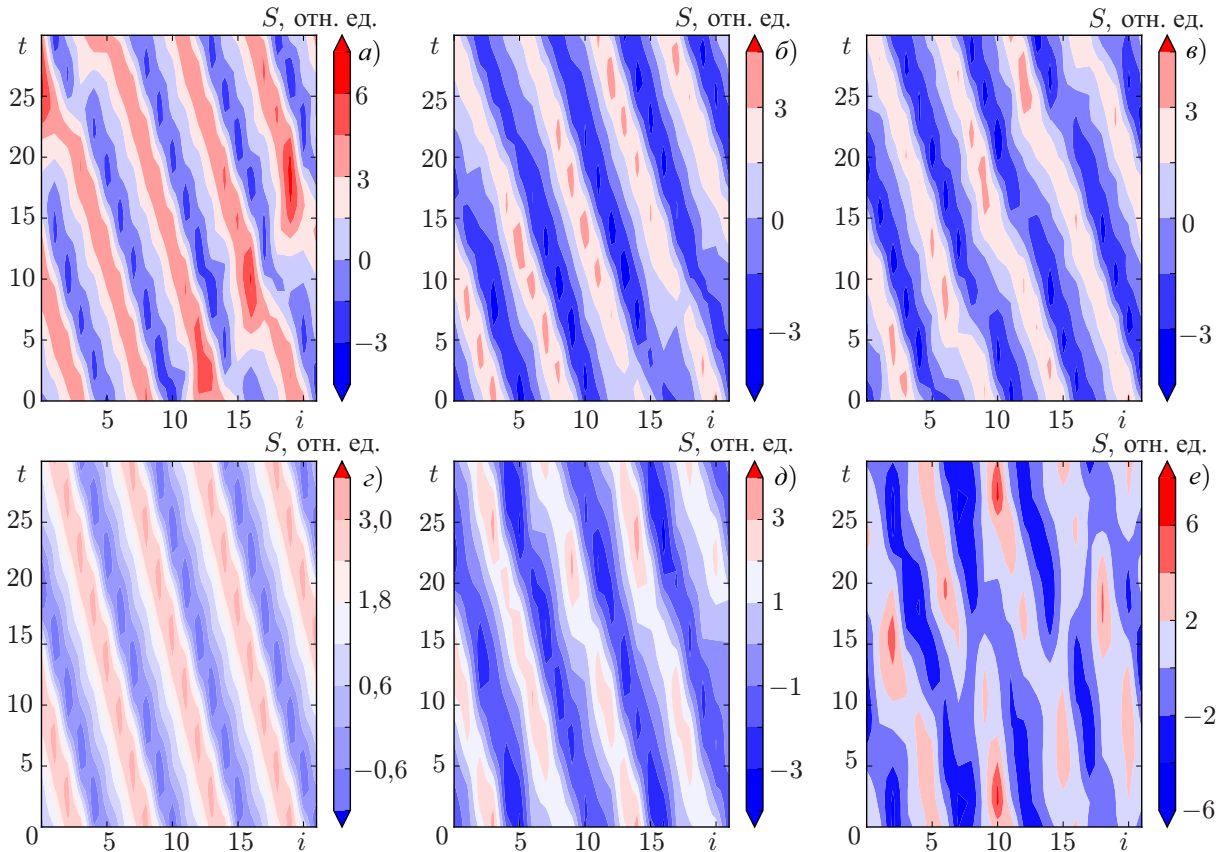


Рис. 5. Пространственно-временная развёртка поля $X_i(t)$ системы Лоренца при $F = 5$ (а), $F = 3$ (г) и пространственно-временная развёртка полей $X_i(t)$ ИНС, соответствующих $F = 5$ (б, в), $F = 3$ (д, е). Структурные параметры ИНС: $m_\varphi = 8, m_Q = 6, m_\psi = 8, d = 5, w = 6$ (б, д) и $m_\varphi = 8, m_Q = 6, m_\psi = 28, d = 5, w = 6$ (в, е)

На рис. 2б, в показано воспроизведение ряда с помощью ИНС при $m_\psi = 8$ (б) и $m_\psi = 28$ (в). Значения остальных параметров ($m_\varphi = 8, m_Q = 6, d = 5, w = 6$) одинаковы для обеих моделей. Среднеквадратичная ошибка аппроксимации данных моделью составляет 0,14 для модели с $m_\psi = 8$ и 0,03 для модели с $m_\psi = 28$.

На рис. 4 изображены проекции аттракторов системы Лоренца и ИНС при $m_\psi = 8$ и $m_\psi = 28$, построенные в пространстве трёх временных рядов системы и моделей соответственно, при $F = 5$, что отвечает началу обучающей выборки, и при $F = 3$ за её пределами. В случае ИНС графики приведены для фиксированного времени t (соответствующего каждому из $F = 3, 5$) в динамическом слое (см. выражение (1)). Дополнительно на рис. 5 показана пространственно-временная развёртка аттракторов (зависимость $X_i(t)$) системы Лоренца и ИНС при $F = 5$ (а–в) и $F = 3$ (г–е), а на рис. 6 представлены спектры переменных X_1 системы Лоренца и ИНС при $F = 5$ (а–в) и при $F = 3$ (г–е). Как видно из рисунков, эмпирические модели качественно воспроизводят динамику и спектральный состав сигнала исследуемой системы как при $F = 5$, соответствующем началу обучающей выборки, так и при $F = 3$, т. е. далеко за её пределами. При этом при большом числе нейронов $m_\psi = 28$ ИНС лучше описывает поведение системы на интервале обучения, в то время как модель с $m_\psi = 8$ обладает лучшей предсказательной силой. По всей видимости, в первом случае мы имеем дело с переобученной моделью, которая «плохо» работает за пределами обучающей выборки. Отметим, что для более обоснованного выбора структурных параметров модели требуется введение меры оптимальности, позволяющей построить модель с до-

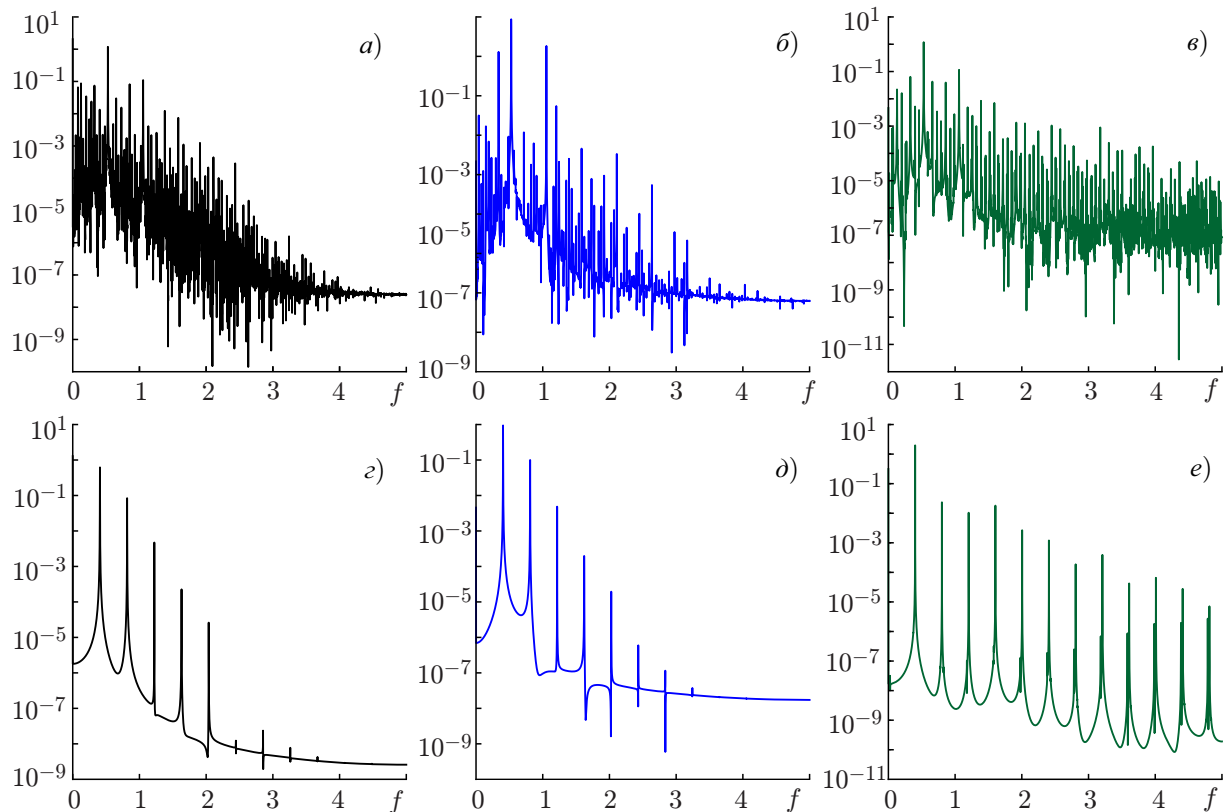


Рис. 6. Спектры рядов переменных X_1 системы Лоренца при $F = 5$ (а), $F = 3$ (з) и спектры рядов ИНС, соответствующие $F = 5$ (б, в), $F = 3$ (д, е). Структурные параметры ИНС: $m_\varphi = 8, m_Q = 6, m_\psi = 8, d = 5, w = 6$ (б, д) и $m_\varphi = 8, m_Q = 6, m_\psi = 28, d = 5, w = 6$ (в, е)

статочной (не избыточной) сложностью. В дальнейшем для этой цели планируется разработать алгоритм расчёта оптимальности ИНС, основанный на байесовой оптимизации, предложенный в работах [34, 35, 54, 71].

Таким образом, в данном разделе было продемонстрировано, что качественные свойства нестационарной динамики системы Лоренца с размерностью фазового пространства $D = 22$ могут быть эффективно описаны эмпирической моделью в форме многослойной рекуррентной ИНС, действующей в пространстве существенно меньшей размерности ($d = 5 \div 7$). При этом явный учёт неавтономности модели позволяет проследить качественное поведение системы и эволюцию её динамического режима не только на интервале обучения, но и за его пределами.

5. ЗАКЛЮЧЕНИЕ

В работе предложена эмпирическая модель в форме рекуррентной ИНС, сочетающей в себе нелинейное отображение наблюдаемых данных в низкоразмерное пространство и реконструкцию оператора эволюции. Такая модель отражает ситуацию, при которой в измеряемых данных с высокой размерностью содержится сигнал некоторой низкоразмерной динамической подсистемы. Предлагаемая ИНС состоит из трёх основных блоков: блок редукции размерности данных (проекция входных векторов на низкоразмерное пространство), блок оператора эволюции, действующего в редуцированном фазовом пространстве, и блок отображения редуцированных фазовых переменных в пространство данных. Процесс обучения ИНС состоит в оптимизации ценовой

функции, получаемой на основе байесова подхода. Фактически, совместная реконструкция фазового пространства и оператора эволюции в рамках единой процедуры обучения даёт фазовые переменные, оптимальные для аппроксимации оператора эволюции.

Работа модели продемонстрирована на примере прогноза динамических переходов в системе Лоренца с медленно меняющимся управляющим параметром по 22-мерному временному ряду. Показана способность модели реконструировать качественное поведение системы как на наблюдаемом участке, так и за его пределами: модель и система имеют аттракторы с одинаковыми характеристиками. В частности, с помощью модели удаётся воспроизвести как корреляционные размерности, так и спектральные характеристики аттракторов системы в широком интервале времени, соответствующем медленному изменению управляющего параметра. Экстраполяция модели за пределы интервала обучения позволяет предсказывать бифуркации, связанные с резким изменением размерности аттрактора, такие как переходы от хаотического режима к квазипериодическому (соответствующему инвариантному тору), и далее к предельному циклу.

В целом, успех в решении задачи эмпирической реконструкции на основе ИНС существенно определяется правильным выбором её структурных параметров, в частности, размерности d редуцированного пространства. В случае реконструкции системы Лоренца диапазон значений d выбирался исходя из известных сведений о примерной величине размерности её аттрактора. Однако при анализе реальных систем в случае общего положения не известны ни размерность самой системы, ни размерности её аттракторов, а имеющиеся временные ряды ограничены по времени. Таким образом, дальнейшее усовершенствование предложенного метода должно быть связано с разработкой эффективного критерия, позволяющего определять оптимальные структурные параметры ИНС при отсутствии каких-либо предварительных сведений об устройстве реконструируемой системы. Первым шагом в этом направлении может стать реализация байесовой обоснованности на базе критериев, разработанных для более простых моделей [54, 71].

Разработка архитектуры рекуррентной искусственной нейронной сети и алгоритма её обучения выполнена при поддержке Российского научного фонда (проект 16–12–10198). Реконструкция и прогноз динамических переходов в системе Лоренца при помощи нейронной сети выполнены при поддержке программы Президиума РАН 51 «Изменение климата: причины, риски, последствия, проблемы адаптации и регулирования» (проект 0035–2018–0024).

СПИСОК ЛИТЕРАТУРЫ

1. Nonlinear dynamics of chaotic and stochastic systems: tutorial and modern developments (Springer Series in Synergetics)/ Eds. V. S. Anishchenko, V. Astakhov, A. Neiman, et al. Berlin, Heidelberg: Springer-Verlag, 2007. 446 p.
2. Gouesbet G., Letellier C. // Physical Rev. E. 1994. V. 49, No. 6. P. 4955.
3. Anishchenko V., Pavlov A., Janson N. // Chaos Solitons & Fractals. 1998. V. 9. P. 1267.
4. Schelter B., Mader M., Mader W., et al. // Europhysics Lett. 2014. V. 105, No. 3. Art. no. 30004.
5. Shandilya S. G., Timme M. // New J. Physics. 2011. V. 13. Art. no. 0130004.
6. Wang W. X., Yang R., Lai Y. C., et al. // Physical Rev. Lett. 2011. V. 106, No. 15. Art. no. 1514101.
7. Baake E., Baake M., Bock H., et al. // Physical Rev. A. 1992. V. 45, No. 8. P. 5524.
8. Bezruchko B. P., Smirnov D., Sysoev I. // Chaos, Solitons & Fractals. 2006. V. 29. P. 82.
9. Gorodetskiy V., Osadchuk M. // Physics Lett., Section A: General, Atomic and Solid State Physics. 2013. V. 377, No. 9. P. 703.

10. Mukhin D. N., Feigin A. M., Loskutov E. M., et al. // *Physical Rev. E. Statistical, Nonlinear, and Soft Matter Physics*. 2006. V. 73, No. 3. Art. no. 036211.
11. Bezruchko B. P., Smirnov D. A. // *Physical Rev. E. Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*. 2001. V. 63, No. 1. Art. no. 016207.
12. Смирнов Д. А., Сысоев И. В., Селезнев Е. П. и др. // *Журн. техн. физ.* 2003. Т. 29, № 19. С. 69.
13. Molkov Y. I., Mukhin D. N., Loskutov E. M., et al. // *Physical Rev. E*. 2011. V. 84, No. 3. Art. no. 036215.
14. Molkov Y. I., Loskutov E. M., Mukhin D. N., et al. // *Physical Rev. E*. 2012. V. 85, No. 3. Art. no. 036216.
15. Ponomarenko V. I., Prokhorov M. D. // *Phys. Rev. E*. 2002. V. 66, No. 2. Art. no. 26215.
16. Sysoev I. V., Prokhorov M. D., Ponomarenko V. I., et al. // *Phys. Rev. E*. 2014. V. 89, No. 6. Art. no. 062911.
17. Sysoev I. V., Ponomarenko V. I., Kulminskiy D. D., et al. // *Phys. Rev. E*. 2016. V. 94, No. 5. Art. no. 052207.
18. Han X., Shen Z., Wang W.-X., et al. // *Phys. Rev. Lett.* 2015. V. 114, No. 2. Art. no. 028701.
19. Brunton S. L., Proctor J. L., Kutz J. N. // *Proc. Natl. Acad. Sci. USA*. 2015. V. 113, No. 15. P. 3932.
20. Mangan N. M., Brunton S. L., Proctor J. L., et al. // *IEEE Trans. Molecular, Biological, and Multi-Scale Communications*. 2016. V. 2, No. 1. P. 52.
21. Pikovsky A. // *Physical Rev. E*. 2016. V. 93, No. 6. Art. no. 062313.
22. Sysoev I. V., Ponomarenko V. I., Pikovsky A. // *Communications in Nonlinear Science and Numerical Simulation*. 2018. V. 57. P. 342.
23. Loskutov E. M., Molkov Y. I., Mukhin D. N., et al. // *Physical Rev. E. Statistical, Nonlinear, and Soft Matter Physics*. 2008. V. 77, No. 6. Art. no. 066214.
24. Jolliffe I. T. *Principal Component Analysis*. 2nd ed. Berlin: Springer, 2002. 487 p.
25. Hannachi A., Jolliffe I. T., Stephenson D. B. // *International J. Climatology*. 2007. V. 27, No. 9. P. 1119.
26. Pires C. A. L., Hannachi A. // *Complexity*. 2017. V. 2017. Art. no. 3076810.
27. Pires C. A. L., Ribeiro A. F. S. // *Climate Dynamics*. 2016. V. 48, No. 3–4. P. 821.
28. M. Ghil, M. R. Allen, M. D. Dettinger, et al. // *Rev. Geophysics*. 2002. V. 40, No. 1. P. 1003.
29. Gámez A. J., Zhou C. S., Timmermann A., et al. // *Nonlinear Processes in Geophys.* 2004. V. 11, No. 3. P. 393.
30. Coifman R. R., Lafon S. // *Applied and Computational Harmonic Analysis*. 2006. V. 21, No. 1. P. 5.
31. Kramer M. A. // *AIChE J.* 1991. V. 37, No. 2. P. 233.
32. Dong D., McAvoy T. // *Computers & Chemical Engineering*. 1996. V. 20, No. 1. P. 65.
33. Tan S., Mayrovouniotis M. L. // *AIChE J.* 1995. V. 41, No. 6. P. 1471.
34. Mukhin D., Gavrilov A., Feigin A., et al. // *Scientific Reports*. 2015. V. 5. P. 15510.
35. Gavrilov A., Mukhin D., Loskutov E., et al. // *Chaos*. 2016. V. 26, No. 12. Art. no. 123101.
36. Kravtsov S., Kondrashov D., Ghil M. // *Stochastic Physics and Climate Modelling*. 2009. No. 1. P. 35.
37. Abarbanel H. D. I. *Analysis of observed chaotic data*. New York: Springer, 1996. 247 p.
38. Bezruchko B. P., Smirnov D. A. // *Springer Series in Synergetics*. 2010. P. 416.
39. Feigin A. M., Molkov Y. I., Mukhin D. N., et al. // *Faraday Discussions*. 2001. No. 120. P. 105.
40. Фейгин А. М., Мо́лков Я. И., Мухин Д. Н. и др. // *Изв. вузов. Радиофизика*. 2001. Т. 44, № 5. С. 376.

41. Mukhin D., Loskutov E., Mukhina A., et al. // J. Climate. 2015. V. 28, No. 5. P. 1940.
42. Mukhin D., Kondrashov D., Loskutov E., et al. // J. Climate. 2015. V. 28, No. 5 P. 1962.
43. de la Iglesia M. D., Tabak E. G. // Communications on Pure and Applied Mathematics. 2013. V. 66, No. 1. P. 48.
44. Bock H. G. // Numerical Treatment of Inverse Problems in Differential and Integral Equations: Proc. Intern. Workshop, Heidelberg, FRG, 30 August–3 September 1982 / Ed. by P. Deuffhard, E. Hairer. Boston: Birkhäuser Boston, 1983. P. 95.
45. Trischler A. P., D'Eleuterio G. M. // Neural Networks. 2016. V. 80. P. 67.
46. Werbos P. J. // Proc. IEEE. 1990. V. 78, No. 10. P. 1550.
47. Feldkamp L. A., Prokhorov D. V., Eagen C. F., et al. // Nonlinear Modeling: Advanced Black-Box Techniques / Ed. by J. A. K. Suykens, J. Vandewalle. Boston, MA: Springer US, 1998. P. 29.
48. Lukoševičius M., Jaeger H. // Computer Science Rev. 2009. V. 3, No. 3. P. 127.
49. Pathak J., Hunt B., Girvan M., et al. // Physical Rev. Lett. 2018. V. 120, No. 2. Art. no. 024102.
50. Lu Z., Hunt B. R., Ott E. // Chaos. 2018. V. 28, No. 6. Art. no. 061104.
51. Williams R. J., Zipser D. // Neural Computation. 1989. V. 1, No. 2. P. 270.
52. Tsung F., Cottrell G. W. // Advances in Neural Information Processing Systems. 1995. V. 7. P. 481.
53. Cybenko G. // Approximation Theory and its Applications. 1989. V. 9, No. 3. P. 17.
54. Gavrilov A., Loskutov E., Mukhin D. // Chaos, Solitons and Fractals. 2017. V. 104. P. 327.
55. Glorot X., Bengio Y. // Proc. Thirteenth Intern. Conf. Artificial Intelligence and Statistics. 13–15 May 2010. Chia Laguna Resort, Sardinia, Italy. 2010. V. 9. P. 249.
56. Hochreiter S. // International J. Uncertainty, Fuzziness and Knowledge-Based Systems. 1998. V. 6, No. 2. P. 107.
57. Kingma D. P., Ba J. // arXiv preprint: arXiv:1412.6980 [cs.LG].
58. Fletcher R. Practical Methods of Optimization. 2nd ed. New York: John Wiley & Sons, 1987. 456 p.
59. Lorenz E. N. // Seminar on Predictability, 4–8 September 1995 / ECMWF. V. 1. Shinfield Park, Reading: ECMWF, 1995. P. 1.
60. Ott E., Hunt B. R., Szunyogh I., et al. // Tellus A: Dynamic Meteorology and Oceanography. 2004. V. 56, No. 5. P. 415.
61. Trevisan A., Palatella L. // Nonlinear Processes in Geophysics. 2011. V. 18, No. 2. P. 243.
62. Crommelin D., Vanden-Eijnden E. // J. Atmospheric Sciences. 2008. V. 65, No. 8. P. 2661.
63. Gallavotti G., Lucarini V. // J. Statistical Physics. 2014. V. 156, No. 6. P. 1027.
64. Pazó D., Szendro I., López J., et al. // Physical Rev. E. 2008. V. 78, No. 1. Art. no. 016209.
65. Lorenz E. N., Emanuel K. A. // J. Atmospheric Sciences. 1998. V. 55, No. 3. P. 399.
66. Karimi A., Paul M. R. // Chaos. 2010. V. 20, No. 4. P. 1.
67. Schuster H. G., Just W. Deterministic chaos: an introduction. 4th ed. Weinheim: Wiley-VCH, 2005. 287 p.
68. Kantz H., Schreiber T. Nonlinear time series analysis. Cambridge: Cambridge University Press, 2003. 388 p.
69. Mañé R. // Lecture Notes in Mathematics. 1981. V. 898. P. 230.
70. Sauer T., Yorke J., Casdagli M. // J. Statistical Physics. 1991. V. 65, No. 3. P. 579.
71. Molkov Y. I., Mukhin D. N., Loskutov E. M., et al. // Phys. Rev. E. 2009. V. 80, No. 4. Art. no. 046207.

Поступила в редакцию 14 июня 2018 г.; принята в печать 25 декабря 2018 г.

**SIMULATION OF COMPLEX SYSTEMS USING THE OBSERVED DATA
BASED ON RECURRENT ARTIFICIAL NEURAL NETWORKS***A. F. Seleznev, A. S. Gavrilov, D. N. Mukhin, E. M. Loskutov, and A. M. Feigin*

We propose a new approach to reconstructing complex spatially distributed systems based on the time series generated by the latter. It allows one to combine two basic steps of such a reconstruction, i.e., the choice of a set of the phase variables of the system using the observed time series and construction of the evolution operator acting in the chosen phase space with the help of artificial network with special topology. This network, first, displays the reference high-dimension data in the lower-dimension space and, second, specifies the evolution operator in this space. The efficiency of this approach is demonstrated by an example of reconstruction of the Lorentz system, i.e., the high-dimension model of atmospheric dynamics.