

# Динамическое физически структурированное моделирование данных в противовес классическому анализу временных рядов

И.В. Семушин<sup>1</sup>, Ю.В. Цыганова<sup>1</sup>

<sup>1</sup>Ульяновский государственный университет, Л. Толстого 42, Ульяновск, Россия, 432017

**Аннотация.** Математическую основу анализа временных рядов составляет подгонка заранее выбранной модели (в простейшем примере - линейной комбинации базисных функций) под экспериментальные данные с одним из критериев качества, например, с критерием наименьших квадратов. Однако какой бы ни была выбранная модель и каким бы ни был критерий качества подгонки, такой подход не предусматривает опоры на понимание тех динамических законов физики - ввиду их сложности или неизвестности, - которые диктуют наблюдаемое поведение данных. Настоящая работа изучает вопрос: какие выгоды или преимущества может приносить математическое моделирование таких законов при его включении в анализ природных или экспериментальных данных.

## 1. Введение

Хорошо известно, что математическую основу анализа временных рядов (ВР) составляет процедура подгонки какой-либо модели под прошлые экспериментальные данные для выявления в них и прогнозирования на будущее четырех главных компонент, называемых: *тренд* – общая тенденция к росту, уменьшению или стагнации на длительном отрезке времени, *цикличность* – наличие повторяющихся сценариев на длительном отрезке времени, *сезонность* – привязанные к временам года и повторяющиеся из года в год картины в изменениях наблюдаемых данных и *иррегулярность* – отсутствие какого-либо особого, заслуживающего внимания, рисунка в поведении данных [1, р. 12]. Тренд понимают как медленно изменяющуюся функцию времени  $m_t$ , цикличность и сезонность часто объединяют в один признак как наличие сезонной (периодической) составляющей  $s_t$ , а иррегулярность трактуют как случайный шум  $Y_t$ . Поэтому классический анализ ВР подразумевает декомпозицию данных на три составляющие [2, р. 23]:

$$X_t = m_t + s_t + Y_t.$$

В анализе ВР разработаны типы (отдельные классы) моделей, из которых производят выбор. В числе типовых моделей выделяют: статические либо динамические, детерминистские либо стохастические, линейные либо нелинейные. Статические модели представляют собой взвешенную комбинацию некоторых базисных функций с последующей подгонкой весовых коэффициентов по критерию близости этой комбинации к имеющимся данным (обычно по критерию наименьших квадратов). В числе линейных динамических стохастических моделей широко известны следующие модели: Autoregressive (AR), Moving Average (MA),

Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Autoregressive Fractionally Integrated Moving Average (ARFIMA) и Seasonal Autoregressive Integrated Moving Average (SARIMA) [1, p. 18]. Их общей особенностью (кроме MA) является то, что текущее значение  $X_t$  выражено как комбинация некоторого числа предыдущих значений  $X_{t-k}$ ,  $k = 1, \dots, n$  наблюдаемых данных (это число  $n$  определяет порядок, или память модели). Кроме этих (вполне традиционных) моделей, предложены искусственные нейронные сети и машины опорных векторов (Support Vector Machines by Vapnik–Chervonenkis) [1, pp. 31–41], но они представляют собой отдельно стоящие альтернативные типы моделей.

Критический этап процедуры анализа ВР – выбор модели, хотя понятие *правильный выбор* заранее не определено. Отбор *адекватной* модели из некоторого класса производят итеративно по схеме, названной *методология Бокса–Дженкинса* [1, pp. 23–24], в три этапа: *идентификация модели, оценивание параметров и диагностическая проверка*. Меры адекватности могут быть разные: Variance Accounted For (VAF), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) или другие. О модели судят постфактум: модель оказалась удачна, целесообразна, уместна, или же нет. Однако такая практика выбора наилучшей модели ВР остается формальной, так как постулирует класс моделей без акцента на те динамические законы физики, – возможно, ввиду их сложности или неизвестности, – которые диктуют наблюдаемое поведение данных. Настоящая работа поднимает вопрос:

*Какие выгоды или преимущества может приносить учет физических представлений при их включении в методологию моделирования данных в качестве ее первого этапа.*

В последние годы замечено, что все традиционные, первоначально возникшие модели ВР могут быть рассмотрены с более общих позиций как специальные случаи моделей в пространстве состояний [2, pp. 259–316]. Этот передовой подход сближает две области знания – анализ ВР и статистически оптимальное оценивание – и тем самым дает возможность воспользоваться хорошо разработанной стохастической теорией предсказания, фильтрации [3, сс. 176–229] и сглаживания [3, сс. 230–280]. Данная работа предлагает искать решение указанного вопроса именно в этом широком классе моделей, но не в общем виде, – что было бы невозможно из-за неисчислимого многообразия изучаемых природных явлений, – а продемонстрировать это на конкретном, реальном примере.

В качестве примера в работе взяты реальные данные суточной термометрии человека, применяемые в медицинской практике, где не столь важно видеть типовые компоненты ВР (тренд, сезонность или шум видны врачу с первого взгляда), а важнее получать более детальные оценки текущего и, главное, будущего состояний пациента, чтобы своевременно предупредить опасное развитие событий.

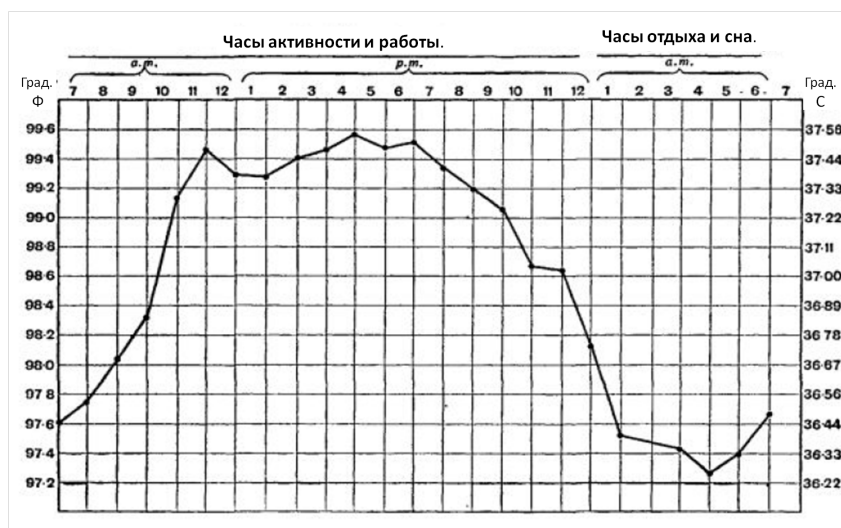
Почему модель в пространстве состояний? – Подгонка модели в пространстве состояний предпочтительна потому, что, как сказано Калманом [4, с. 13], “нужно потребовать, чтобы множество внутренних состояний системы было достаточно богатым, чтобы вместить всю информацию о предыстории системы, необходимой для предсказания влияния прошлого на будущее”.

Далее текст статьи организован следующим образом. В разделе 2 характеризованы реальные данные – результаты суточного мониторинга температуры тела здорового человека из базы данных конкретной больницы. Раздел 3 содержит представление об этих данных как о результате наблюдения за работой скрытого от нас механизма теплового гомеостаза – в виде дискретных уравнений в пространстве состояний с последующей идентификацией неизвестных параметров. Следующие разделы 4 и 5 введены для сравнения: один альтернативный подход в разделе 4 основан на формальной подгонке неизвестных коэффициентов линейной модели измерительных данных по схеме классического метода наименьших квадратов (МНК); другой – в разделе 5 – на аппроксимации измерительных данных полиномом  $n$ -ой степени, при этом искомая степень определяется из условия минимума нормы невязки аппроксимации, а для придания

стохастичности измерениям в полиномиальную модель добавлен гауссовский шум с распределением  $\mathcal{N}(0, 0.01)$ .

## 2. Суточная термометрия процесса теплового гомеостаза человека

Механизм гомеостаза скрыт от нас, но проявляется в процессе терморегуляции живого организма. Рассмотрим один из многочисленных примеров графика суточной динамики показателей температуры тела здорового человека (рисунок 1). Длительное наблюдение



**Рисунок 1.** График суточной термометрии человека (a.m. / p.m. – до / после полудня).  
 Источник: ANTRANIK.org <http://antranik.org/regulation-of-body-temperature/>.

за этим процессом у множества людей обнаруживает наличие в нем периодической и стохастической составляющих [5].

Подбирая для них модель, следуем *принципу экономии* – он известен как *принцип бритвы Оккама* [1, р. 16]: “сталкиваясь с рядом конкурирующих и адекватных объяснений, выбирайте наиболее простое”. Наиболее простым генератором периодической компоненты является гармонический осциллятор  $x_1(t_k) = A_N \sin(k\omega_N \tau + \phi_N)$  с неопределенными, в общем случае, амплитудой  $A_N$  и фазой  $\phi_N$ , но с известным периодом (сутки, 24 часа). Неопределенность  $A_N$  задает возможное наличие аномалий, то есть отклонений от понятия “здоровый человек”, а неопределенность  $\phi_N$  допускает смещение ритма вдоль оси времени.

Стохастическую составляющую модели теплового гомеостаза представим гауссовским марковским процессом первого порядка с двумя неизвестными параметрами, которые требуют идентификации: диффузия и интервал корреляции (коэффициент сноса).

Основаниями для этих физических представлений служат реальные данные суточной термометрии<sup>1</sup>. Суточная термометрия выполнялась по специальному плану эксперимента [8]. В его реализации принимали участие здоровые добровольцы среднего (почти одинакового) возраста. Согласно плану, замеры температуры производились датчиками, наклеенными на теле, каждые 5 минут с точностью  $\pm 0.1^\circ\text{C}$  в течение суток. Для проведения топической суточной термометрии применяли программно-аппаратный комплекс КМТП-01-МИДА и соответствующую оригинальную методику (Патент РФ № 2344750, приоритет

<sup>1</sup> Данные термометрии любезно предоставил А.Б. Песков – доктор медицинских наук, профессор Ульяновского государственного университета.

от 13.06.2007). Были получены реализации процесса циркадианного ритма температуры тела группы здоровых людей. Их осреднение приводит к понятию “нормальный ритм”.

### 3. Динамическое физически структурированное моделирование данных

Основным инструментом изучения сложных биологических процессов – таких как гомеостаз (homeostasis) [6] – является построение для них и последующая идентификация адекватной математической модели. Такие исследования активно проводятся [7]. Линейные дискретные динамические модели суточной динамики теплового гомеостаза, представленные разностными уравнениями в пространстве состояний, впервые построены в [9]. Метод активной параметрической идентификации таких моделей подробно рассмотрен в [10]. Следуя этим работам, представим динамическую модель суточной термометрии теплового гомеостаза здорового человека в следующем виде:

$$\left. \begin{aligned} dx_1(t) &= x_2(t)dt \\ dx_2(t) &= -\omega_N^2 x_1(t)dt \\ dx_3(t) &= -(1/T)(x_3(t) - \theta^*)dt + \sigma\sqrt{2/T}d\dot{\beta}(t) \end{aligned} \right\} \begin{array}{l} \text{уравнения} \\ \text{состояния} \\ \text{физической} \\ \text{модели} \end{array} \quad (1)$$

где  $\omega_N = 2\pi/T_N$  rad/min,  $T_N = 24$  h,  $x_1(t_k) = a_N \sin(k\omega_N\tau) + b_N \cos(k\omega_N\tau)$ ,  $x_3(t)$  – случайный процесс Орнштейна-Уленбека, параметры  $T > 0$  и  $\sigma > 0$  – неизвестны,  $\dot{\beta}(t)$  – стандартный винеровский процесс (с нулевым средним и единичной диффузией),  $\lim_{t_0 \rightarrow -\infty} \dot{\beta}(t_0) = 0$  (a.s.),  $\theta^*$  – среднесуточное значение температуры тела;  $a_N$  и  $b_N$  – параметры амплитуды  $A_N$  синусоидальной компоненты  $x_1(t_k)$  текущей температуры тела:  $A_N = \sqrt{a_N^2 + b_N^2}$ .

Имеем: текущая температура  $x_1(t) + x_3(t)$  измеряется каждые  $\tau = 5$  минут датчиком с погрешностью измерения  $v(t_k)$ ; процесс  $v \triangleq \{v(t_k)\}$  моделируется как дискретный белый шум с ковариацией  $R$ :

$$y(t_k) = x_1(t_k) + x_3(t_k) + v(t_k) \quad \left. \right\} \begin{array}{l} \text{уравнение} \\ \text{измерений} \end{array} \quad (2)$$

Для системы (1)–(2) предложены различные эквивалентные представления, получаемые невырожденными преобразованиями базиса [9]. Используем один из этих вариантов – модель **3dDRCM** (3-мерная жорданова модель в вещественном базисе, помеченном  $\star$ ):

$$y_k = \underbrace{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}}_H x_k^* + v_k, \quad k = 1, 2, \dots, K \quad (3)$$

$$\begin{aligned} x_{k+1}^* &= \underbrace{\begin{bmatrix} \cos \omega_N \tau & -\sin \omega_N \tau & 0 \\ \sin \omega_N \tau & \cos \omega_N \tau & 0 \\ 0 & 0 & d \end{bmatrix}}_\Phi x_k^* + \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1-d \end{bmatrix}}_{B_d} u_k \\ &+ \underbrace{\begin{bmatrix} 0 \\ 0 \\ \sigma\sqrt{1-d^2} \end{bmatrix}}_{G_d} w_{d,k}, \quad k = 0, 1, \dots, \quad x_0 = 0.65 [0.5 \mid -0.5 \mid 0.0]^T \end{aligned} \quad (4)$$

где  $\lambda \triangleq 1/T$ ,  $d \triangleq e^{-\lambda\tau}$  – неизвестные параметры. В модели (3)–(4)  $w_{d,k}$  – дискретный белый шум с нулевым средним и ковариацией  $Q = 1.0$ ;  $u_k = \theta^*$ ; ошибка измерений

$v_k \sim \mathcal{N}(0, R)$  с ковариацией  $R = (0.1)^2$ . На данном этапе считаем, что неопределенный векторный параметр модели  $\theta \triangleq [\lambda \mid \sigma]^T$  состоит из двух неизвестных величин:  $\lambda$  и  $\sigma$ . Требуется идентифицировать “истинное” значение  $\hat{\theta} = [\hat{\lambda} \mid \hat{\sigma}]^T$  этого параметра по имеющимся измерительным данным  $Y_1^K = \{y_1, y_2, \dots, y_K\}$ , где  $K$  — количество измерений;  $K = 288$  для суточного мониторинга с интервалом в 5 минут. Под “истинным” понимаем то значение, которое обеспечивает наилучшую аппроксимацию сложного процесса гомеостаза суточной температуры для в среднем здорового человека в классе линейных инвариантных во времени систем, которому принадлежит физически структурированная модель (3)–(4).

**Первый этап** – (1°) идентификация векторного параметра  $\theta$  модели (3)–(4) по каждой реализации отдельно и (2°) проверка адекватности этой модели, когда она использует среднее  $\bar{\theta}$  из отдельно полученных оценочных значений  $\hat{\theta}_n$  параметра  $\theta$ ,  $n = 1, \dots, N$ ,  $N = 18$  (выборка температурных кривых с одинаковым временем начала измерений – 19:30 – в течение одних суток). Вычислительные эксперименты 1° и 2° первого этапа проводим на имеющихся данных (см. раздел 2).

*Эксперимент 1°*: По каждой реализации проведем параметрическую идентификацию. Для этого применяем метод вспомогательного функционала качества (ВФК) [11, 12, 13] с минимизацией ВФК одним из известных градиентных методов [14] – используем МАТЛАБ-функцию `fminunc`. Отдельные результаты осредним:  $\bar{\theta} = \sum_{n=1}^N \hat{\theta}_n$ . Получены следующие средние оценочные значения:  $\bar{\lambda} = 0.0168$  и  $\bar{\sigma} = 1.9252$ .

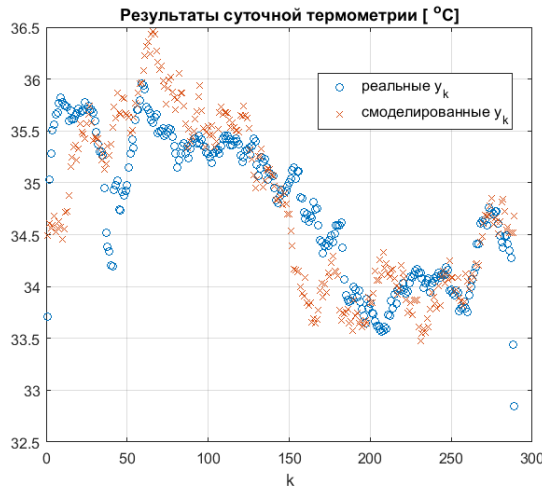
*Эксперимент 2°*: Эти значения  $\bar{\lambda}$  и  $\bar{\sigma}$  подставим в модель **3dDRCM** и сгенерируем на их основе  $N = 18$  реализаций модельных измерений, чтобы их осреднить и затем сравнить с реальными осредненными измерениями с целью подтвердить или опровергнуть правомочность применения предложенной модели (3)–(4) с этими оцененными  $\bar{\lambda}$  и  $\bar{\sigma}$ ; результат – ниже на рисунке 2. Средние по интервалу измерений значения для осредненных реальных и модельных данных равны, соответственно,  $m_1 = 34.78$  и  $m_2 = 34.73$ . Сниженный средний уровень температуры объясняется тем, что в некоторых температурных кривых из рассмотренной выборки присутствуют аномальные данные, полученные вследствие отклеивания от кожного покрова таблетки-термометра.

Визуально можно судить, что модельные данные, в целом, повторяют картину реальных данных. Следовательно, данная модель, являясь простой, может быть принята, поскольку “схватывает” эффект скрытого от нас процесса теплового гомеостаза, а не просто аппроксимирует данные. На этом основании считаем, что теперь можно принять оценочные значения параметров  $\bar{\lambda} = 0.0168$  и  $\bar{\sigma} = 1.9252$  за “истинные”:  $\hat{\lambda} = 0.0168$  и  $\hat{\sigma} = 1.9252$ .

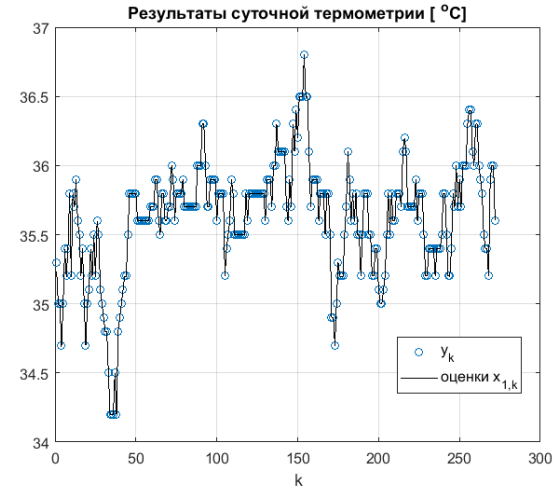
**Второй этап** – испытания возможностей построенного оптимального оценителя. Найденные значения  $\hat{\lambda}$  и  $\hat{\sigma}$  подставим в уравнения фильтра Калмана, который соответствует модели **3dDRCM**, и с его помощью будем оценивать температурное состояние случайно выбранного пациента по его реальным данным; результат – на рисунке 3.

По рисунку 3 видно, что оценки температуры согласуются с реальными данными пациента: этот метод позволяет получать удовлетворительные оценки температуры тела человека по измерениям прошедшего дня и, возможно, предсказывать ход процесса на следующий день.

Результаты вычислительных экспериментов этого раздела, обобщенно представленные рисунками 2 и 3, подтверждают целесообразность применения изложенного подхода.



**Рисунок 2.** Реальные данные измерений (синие кружки) и измерения, смоделированные на основе модели **3dDRCM** (красные крестики). Оба графика – после усреднения по  $N = 18$  реализациям.



**Рисунок 3.** Реальные данные измерений (синие кружки) и оценки температуры, полученные при вычислении по алгоритму Калмана, построенного на основе модели **3dDRCM** (черная сплошная линия).

#### 4. Классическая аппроксимация данных по методу наименьших квадратов

Биофизический закон циркадианного ритма температуры тела человека (выше рисунок 1), в строгом виде не известен, и это вынуждает представлять лишь результат его действия. Гармонический осциллятор – первые два уравнения в (1) – порождает синусоидальный процесс  $x_1(t_k)$ . Анализ ВР предлагает аппроксимировать этой готовой функцией данные  $y_k \triangleq y(k\tau)$ , полученные в моменты  $t_k \triangleq k\tau$  ( $k = 1, \dots, K$ ), то есть полагать приближенно

$$\begin{aligned}
 y_k &\approx A_N \sin(k\omega_N\tau + \phi_N) + \theta^* = a_N \sin(k\omega_N\tau) + b_N \cos(k\omega_N\tau) + \theta^* = x_1(t_k) \\
 v_k &\triangleq y_k - \tilde{y}_k; \quad \tilde{y}_k \triangleq a_N \sin(k\omega_N\tau) + b_N \cos(k\omega_N\tau) + \theta^* \\
 k &= 1, 2, \dots, K > 2, \\
 a_N &= A_N \cos \phi_N, \quad b_N = A_N \sin \phi_N
 \end{aligned}
 \tag{5}$$

Таким образом, данный вектор  $y \triangleq [y_1 \mid \dots \mid y_K]^T$  моделируют произведением матрицы

$$H = \left[ \begin{array}{c|c} \sin(1\omega_N\tau) & \cos(1\omega_N\tau) \\ \sin(2\omega_N\tau) & \cos(2\omega_N\tau) \\ \dots & \dots \\ \sin(K\omega_N\tau) & \cos(K\omega_N\tau) \end{array} \right]
 \tag{6}$$

и неизвестного двумерного вектора  $x = [\bar{x}_1 \mid \bar{x}_2]^T$  параметров  $\bar{x}_1 \triangleq a_N$  и  $\bar{x}_2 \triangleq b_N$ :  $\tilde{y}_k \triangleq h_k^T x + \theta^*$ .

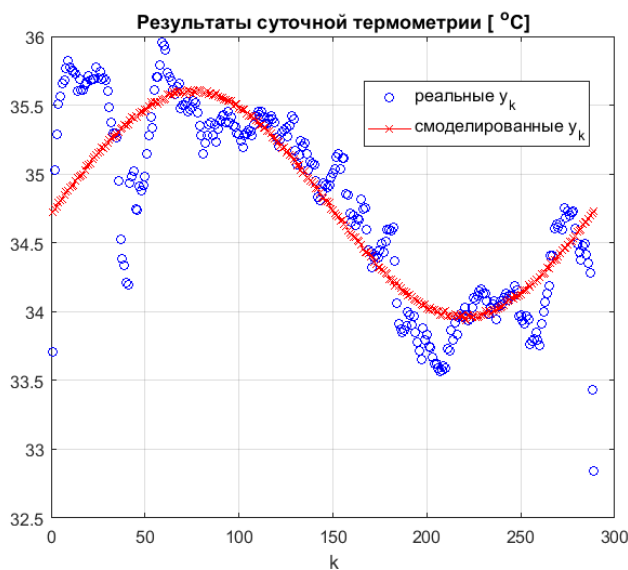
Такая модель является линейной: в нашем случае  $\omega_N \triangleq 2\pi/T_N$  известна, так как период  $T_N$  составляет 24 часа (1440 минут). В других, более сложных задачах частота  $\omega_N \triangleq 2\pi/T_N$  может оказаться неизвестна, что сделает эту модель существенно нелинейной.

Принимая  $\tilde{y} \triangleq Nx + \theta^* \mathbf{1}$  в (5) за модель данных, осуществим подгонку параметров модели по методу наименьших квадратов к вектору  $y$ , беря в качестве  $y$  *осредненный* по множеству реализаций вектор данных. МНК приводит к отысканию наилучшей оценки  $\hat{x}$  вектора  $x$  из системы нормальных уравнений

$$(H^T H) \hat{x} = H^T (y - \theta^* \mathbf{1}). \tag{7}$$

Таким образом, в эксперименте используем реальные данные суточной термометрии из раздела 2, но после их осреднения по  $N = 18$  реализациям. В результате вычислений получаем: уравнение (7) с матрицей (6) имеет решение  $\hat{x}$  с элементами:  $\hat{x}_1 = a_N = 0.8161$  и  $\hat{x}_2 = b_N = -0.0661$ .

Построим графики осредненной реализации  $y_k$  и оценочных модельных измерений  $\hat{y}_k = h_k^T \hat{x} + \theta^*$  (рисунок 4). В результате имеем приближение реальных данных синусоидой, – результат вполне ожидаемый. Извлекать из данных полезную адекватную модель этим методом не удастся. Да, МНК позволяет строить модель, которая подтверждает наличие периодической компоненты (заранее известный факт), позволяет оценивать амплитуду и фазу этой компоненты (что само по себе полезно), но более детальной картины для *отдельно взятой* реализации данных этот метод не выявляет, – происходит грубое осреднение.



**Рисунок 4.** Результаты МНК-моделирования осредненной реализации данных.

*В любом случае:* принципиальное отличие методики этого раздела от методики раздела 3 заключается в том, что здесь внутренний динамический закон, порождающий данные процесса термометрии, – например, в виде уравнений (1) и (2), – полностью проигнорирован.

### 5. Аппроксимация полиномиальной функцией

Возможности МНК для аппроксимации данных суточной термометрии подходом из раздела 4 не исчерпываются. Протестируем их на классической модели в виде полинома степени  $n$ . Для набора данных  $(x_i, y_i), i = 1, \dots, K$  потребуем найти полином  $n$ -ой степени

$$p^{(n)}(x) = p_1 x^n + p_2 x^{n-1} + \dots + p_n x + p_{n+1},$$

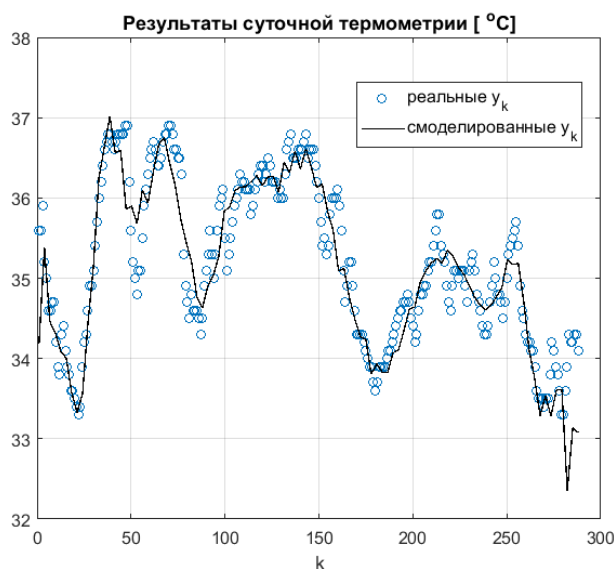
коэффициенты которого являются решением задачи многомерной оптимизации

$$\min_{p_1, p_2, \dots, p_{n+1}} \sum_{i=1}^K (p^{(n)}(x_i) - y_i)^2.$$

Степень полинома должна быть меньше количества заданных точек для того, чтобы такой полином был единственным, т. е.  $n < K$ , и в целях сглаживания.

Вычислительные эксперименты проведем на тех же реальных данных суточной термометрии. Поиск коэффициентов полинома выполним с помощью МАТЛАБ-функции `polyfit`, степень полинома определяем из условия минимума нормы невязки

$$normr = \sqrt{\min_{p_1, p_2, \dots, p_{n+1}} \sum_{i=1}^K (p^{(n)}(x_i) - y_i)^2}.$$



**Рисунок 5.** Реальные данные измерений (синие кружки) и смоделированные данные на основе полинома 30-й степени измерения.

Рисунок 5 показывает реальные данные измерений  $y_k$  (случайно выбранного пациента) и смоделированные измерения на основе полинома 30-й степени. По рисунку 5 видно, что модельные данные хорошо приближают прошлые реальные данные. Это может привести к выводу, что данный подход, якобы, целесообразно использовать для моделирования процесса суточной термометрии человека. На самом деле это ошибочный вывод. Высокой степенью полинома можно добиваться того, что полином пройдет практически через все точки графика реальных данных. Однако при этом исчезает главная цель: выявлять на имеющихся данных типичные тенденции, общие закономерности или скрытые физические механизмы, присущие явлению как таковому, а не отдельно взятой реализации процесса.

## 6. Заключение

Общезначимые выводы данной работы базируются на результатах трех различных подходов к моделированию временных рядов на примере задачи анализа экспериментальных данных суточной термометрии группы здоровых добровольцев среднего возраста.



Первый подход к моделированию данных суточной термометрии использует представление об этих данных как о наблюдаемых результатах работы скрытого биофизического механизма теплового гомеостаза человеческого организма. Чтобы учесть динамический характер этого механизма, его модель выражена стохастическими рекуррентными уравнениями в пространстве состояний с алгоритмом последующей идентификации модельных параметров. Численные эксперименты подтверждают, что включение такой (динамической) модели порождения данных придает алгоритму способность “схватывать” динамическую природу механизма теплового гомеостаза, а не просто аппроксимировать прошлые экспериментальные данные. Эта особенность особо полезна для прогнозирования процесса на ближайшее будущее, например, на следующие сутки.

В противоположность этому, второй рассмотренный подход к моделированию данных суточной термометрии выполнен по классической формальной схеме МНК-подгонки параметров линейной статической модели к имеющимся реальным данным. Результаты экспериментов показывают, что в данном случае получить полезную модель не удастся. Метод наименьших квадратов в этой схеме позволяет строить параметризованную форму функции, удовлетворяющей уравнению детерминистской части той динамической модели, которая применена в первом подходе, но он не восстанавливает саму эту динамическую модель. Численный эксперимент подтверждает априорно известный периодический ход процесса. Само по себе это дает лишь грубую аппроксимацию прошлых данных. Для прогнозирования процесса такой результат большого интереса не представляет.

Третий альтернативный подход основан на аппроксимации измерительных данных полиномом  $n$ -ой степени, при этом искомая степень определяется из условия минимума нормы невязки аппроксимации, а для придания стохастичности модельным измерениям в их полиномиальную модель добавлен дискретный белый гауссовский шум. Хотя в этом случае модельные данные хорошо приближают реальные данные, но это достигается формальным увеличением степени полинома. Данный подход не рекомендуется применять. Очевидная причина заключается в том, что при таком подходе исчезает главная цель моделирования: выявлять из имеющихся данных скрытые физические механизмы, присущие явлению как таковому, а не отдельно взятой реализации данных.

В целом, результаты проведенного сравнительного экспериментального исследования показывают, что динамическое физически структурированное моделирование реальных данных (первый рассмотренный подход) производится не посредством подгонки параметров формально выбранного модельного уравнения в рамках стандартной методологии анализа временных рядов, а благодаря обоснованному включению в модель тех динамических (дифференциальных или разностных) моделей состояния физических процессов, которые порождают наблюдаемое поведение данных. Это позволяет отойти от формализмов привычной методологии и тем добиваться большей точности и полезности результатов анализа, оценивания и прогнозирования физических процессов в сложных системах.

## 7. Литература

- [1] Adhikari, R. An Introductory Study on Time Series Modeling and Forecasting / R. Adhikari, R.K. Agrawal. – LAP Lambert Academic Publishing, arXiv: 1302.6613, 2013. – 67 p.
- [2] Brockwell, P.J. Introduction to Time Series and Forecasting / P.J. Brockwell, R.A. Davis. – New York : Springer-Verlag, Inc., 2002. – 449 p.
- [3] Медич, Дж. Статистически оптимальные линейные оценки и управление. – М.: Энергия, 1973. – 440 с.
- [4] Kalman, R.E. Topics in Mathematical System Theory / R.E. Kalman, P.L. Falb, M.A. Arbib // New York, USA: Mc. Graw Hill, 1969.
- [5] Kelly, G. Body temperature variability (Part 1): a review of the history of body temperature and its variability due to site selection, biological rhythms, fitness, and aging // Altern. Med. Rev. – 2006. – Vol. 11(4). – P. 278-93.

- [6] Нефедов, В.П. Гомеостаз на различных уровнях организации биосистем / В.П. Нефедов, А.А. Ясайтис, В.Н. Новосельцев. – Новосибирск: Наука. Сиб. отд-ние, 1991. – 232 с.
- [7] Kingma, B.R.M. Human Thermoregulation, A synergy between physiology and mathematical modelling: Ph.D. thesis // School of Engineering. – Universiteit Maastricht, the Netherlands, 2012.
- [8] Майорова, Е.А. Характеристики суточных температурных кривых, полученных с различных участков кожных покровов здорового человека / Е.А. Майорова, А.Б. Песков, М.П. Хохлов // MEDLINE.RU. Российский биомедицинский журнал. – 2014. – Т. 15, № 3. – С. 629-638.
- [9] Semushin, I.V. Identification of a simple homeostasis stochastic model based on active principle of adaptation / I.V. Semushin, J.V. Tsyganova, A.G. Skovikov // Proceedings of International Conference "Applied Stochastic Models and Data Analysis ASMDA & DEMOGRAPHICS. – Spain: Barcelona, 2013. – P. 775-783.
- [10] Semushin, I.V. Identification of human body daily temperature dynamics via minimum state prediction error method / I.V. Semushin, J.V. Tsyganova, M.V. Kulikova // Proceedings of ECC, European Control Conference, 2016. – P. 2429-2434.
- [11] Семушин, И.В. Идентификация линейных стохастических объектов по неполным зашумленным измерениям вектора состояния // Автоматика и телемеханика. – 1985. – № 8. – С. 61-71.
- [12] Semushin, I.V. Adaptation in stochastic dynamic systems - survey and new results II // Int. J. Communications, Network and System Sciences. – 2011. – Vol. 4(4). – P. 266-285.
- [13] Semushin, I.V. Adaptation in stochastic dynamic systems – survey and new results IV: Seeking minimum of API in parameters of data / I.V. Semushin, J.V. Tsyganova // Int. J. Communications, Network, and System Sciences. – 2013. – Vol. 6(12). – P. 513-518.
- [14] Васильев, Ф.П. Методы оптимизации. – М.: Факториал Пресс, 2002. – 824 с.

### Благодарности

Авторы благодарят доктора медицинских наук, профессора Ульяновского государственного университета Андрея Борисовича Пескова за предоставленные реальные данные суточной термометрии из базы данных Ульяновской областной больницы.

Работа выполнена при финансовой поддержке РФФИ и Правительства Ульяновской области в рамках научного проекта № 18-47-730001 p\_a.

## Dynamical physically structured data modeling vs. classical time series analysis

I.V. Semushin<sup>1</sup>, Yu.V. Tsyganova<sup>1</sup>

<sup>1</sup>Ulyanovsk State University, L. Tolstoy street 42, Ulyanovsk, Russia, 432017

**Abstract.** The mathematical groundwork for the time series analysis is the process of fitting models to data and analyzing the accuracy of the fit. Linear models are popular due to their flexibility to represent several varieties of time series with simplicity by a linear combination of some basis functions. One of the frequently used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA) model, which has subclasses of other models. Such wellknown models have one general property: they do not include an understanding of those dynamic laws of physics - because of their complexity or uncertainty - that dictate the observed behavior of the data. This paper addresses the following question: What benefits or advantages can we obtain due to mathematical modeling of such laws when we include them in the data analysis.