

СУБЪЕКТИВНОЕ ВОССТАНОВЛЕНИЕ ПРОПУСКОВ ДАННЫХ ИЗМЕРЕНИЙ ОБЪЕКТА ИССЛЕДОВАНИЯ И ЕГО МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Пытьев Ю.П., Чуличков А.И., Фаломкина О.В.
МГУ им. М.В. Ломоносова, кафедра математического
моделирования и информатики
[E-mail: yuri.pytyev@gmail.com](mailto:yuri.pytyev@gmail.com), achulichkov@gmail.com,
olesya.falomkina@gmail.com

МАТЕМАТИЧЕСКИЙ ФОРМАЛИЗМ СУБЪЕКТИВНОГО МОДЕЛИРОВАНИЯ

Основа МФСМ — пространство $(X, P(X), Pl^{\tilde{x}}, Bel^{\tilde{x}})$, $P(X)$ — класс всех подмножеств X , $Pl^{\tilde{x}}(\cdot): P(X) \rightarrow L$ и $Bel^{\tilde{x}}(\cdot): P(X) \rightarrow \hat{L}$, где $L = ([0,1], \leq, +, \times) = ([0,1], \leq, \max, \min)$ и $\hat{L} = ([\hat{0}, \hat{1}], \hat{\leq}, \hat{+}, \hat{\times}) = ([0,1], \geq, \min, \max)$ суть шкалы значений Pl и Bel . Меры правдоподобия и доверия заданы м.-и. равенствами: $\forall E \in P(X)$

$$Pl_X(E) \equiv Pl^{\tilde{x}}(E) = Pl^{\tilde{x}}(\tilde{x} \in E) = \underset{x \in E}{+} t^{\tilde{x}}(x) =$$

$$= \sup_{x \in E} t^{\tilde{x}}(x), \quad E \neq \emptyset, \quad Pl^{\tilde{x}}(\emptyset) \stackrel{\text{def}}{=} 0, \quad Pl^{\tilde{x}}(X) \stackrel{\text{def}}{=} 1, \quad (*)$$

$$Bel_X(E) \equiv Bel^{\tilde{x}}(E) = Bel^{\tilde{x}}(\tilde{x} \in E) = \underset{x \in X \setminus E}{\hat{+}} \hat{t}^{\tilde{x}}(x) =$$

$$= \inf_{x \in X \setminus E} \hat{t}^{\tilde{x}}(x), \quad E \neq X, \quad Bel^{\tilde{x}}(X) \stackrel{\text{def}}{=} 1, \quad Bel^{\tilde{x}}(\emptyset) \stackrel{\text{def}}{=} 0,$$

в которых $E = \bigcup_{x \in E} \{x\} = \bigcap_{x \in X \setminus E} (X \setminus \{x\})$,

$$t^{\tilde{x}}(x) \stackrel{\text{def}}{=} Pl^{\tilde{x}}(\tilde{x} = x), \quad \hat{t}^{\tilde{x}}(x) \stackrel{\text{def}}{=} Bel^{\tilde{x}}(\tilde{x} \neq x), \quad x \in X. \quad (**)$$

СУБЪЕКТИВНОЕ МОДЕЛИРОВАНИЕ ДАННЫХ ИЗМЕРЕНИЙ И КРИТЕРИИ ОПТИМАЛЬНОСТИ СУБЪЕКТИВНЫХ РЕШЕНИЙ

Субъективная модель измерений, определённая модельером-исследователем, имеет вид

$$y_i = f(t_i) + v_i, \quad i = 1, \dots, n, \quad (3)$$

где y_i — искажённое шумом v_i значение температуры воды $f(t_i)$ в момент времени $t_i, i = 1, \dots, n$, $f(\cdot): [a, b] \rightarrow \mathcal{R}^1$ — гладкая функция, квадрат которой и её производных интегрируемы, $v_i, i = 1, \dots, n$, — ошибки измерений, шум.

1. В силу физических свойств воды её температура $f(\cdot)$ под действием внешних факторов изменяется плавно, без резких скачков.
2. Шум $v_i, i = 1, \dots, n$, считается случайным, математическая модель его неизвестна.

СУБЪЕКТИВНОЕ МОДЕЛИРОВАНИЕ ДАННЫХ ИЗМЕРЕНИЙ И КРИТЕРИИ ОПТИМАЛЬНОСТИ СУБЪЕКТИВНЫХ РЕШЕНИЙ

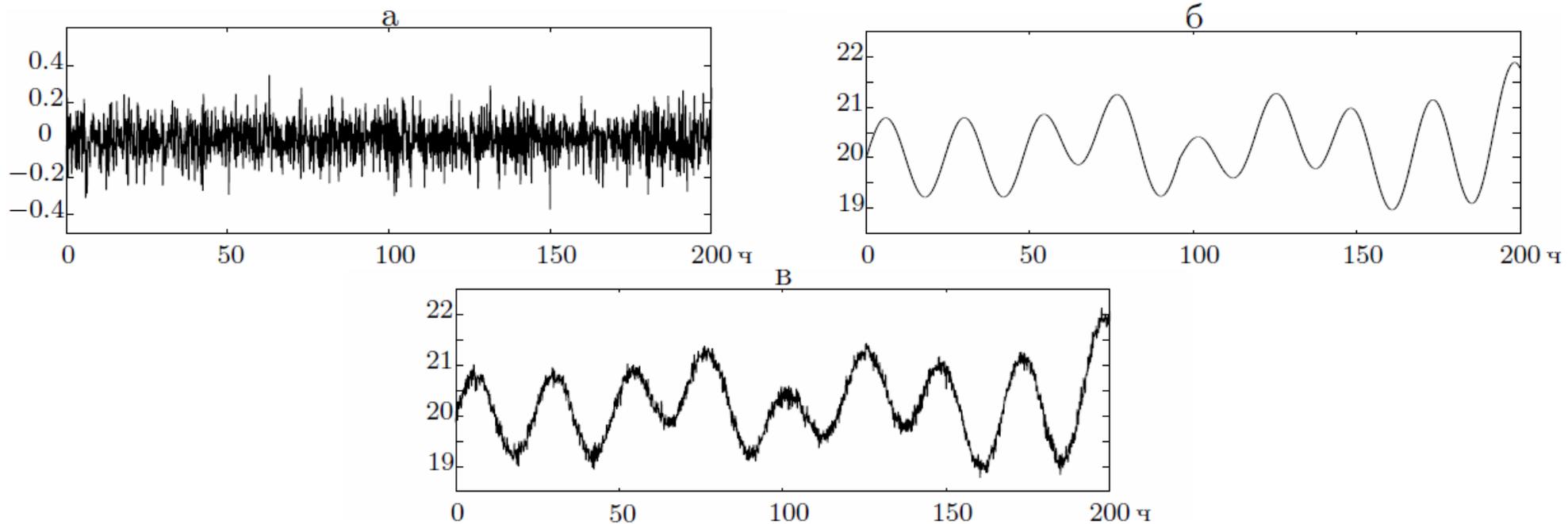


Рис. 1. а) График шума $\nu(t_i)$, $i = 1, \dots, n$. б) Значения температуры воды $f(t_i)$, $i = 1, \dots, n$. в)

Данные измерений температуры как функции времени $y_i = f(t_i) + \nu_i$, $i = 1, \dots, n \approx 1500$.

СГЛАЖИВАЮЩИЕ СПЛАЙНЫ

Сплайном порядка q называется функция $s(t)$, $t \in [a, b]$, удовлетворяющая следующим условиям:

- для фиксированных t_1, \dots, t_n , $a < t_1 < \dots < t_n < b$, называемых узлами сплайна, $s(\cdot)$ – полином степени, меньшей или равной $2q - 1$ на интервалах (t_i, t_{i+1}) , $i = 1, \dots, n - 1$, где $n \geq q$;
- $s(\cdot)$ – полином степени, меньшей или равной $q - 1$ на $[a, t_1) \cup (t_n, b]$;
- производная порядка $2q - 2$ $s^{(2q-2)}(\cdot) \in C_{[a,b]}$, где $C_{[a,b]}$ – класс непрерывных на $[a, b]$ функций, т. е. в узлах t_1, \dots, t_n полиномы сшиты по непрерывности вплоть до производной порядка $2q - 2$: $0 = s^{(2q-2)}(t_j + 0) - s^{(2q-2)}(t_j - 0)$.

СГЛАЖИВАЮЩИЕ СПЛАЙНЫ

Сглаживающий сплайн *порядка* q определим как решение $f(\cdot) = s(\cdot)$ следующей задачи на минимум:

$$\min\{\rho \int_a^b (f^{(q)}(t))^2 dt + \sum_{j=1}^n (f(t_j) - y_j)^2 \mid f \in W_2^q\}, \quad (1)$$

его гладкость тем выше, чем больше порядок сплайна q *и чем больше сглаживающий фактор* ρ , гильбертово пространство W_2^q , в котором решается задача на минимум (1), определено скалярным произведением

$$(f, g) = \rho \int_a^b f^{(q)}(t)g^{(q)}(t) dt + \sum_{i=1}^n f(t_i)g(t_i), \quad (2)$$

в котором $(q - 1)$ -е производные функций f и g абсолютно непрерывны на $[a, b]$, а q -е производные принадлежат гильбертову пространству $L_2[a, b]$.

При $n \geq q$ задача (1) имеет единственное решение при любых $y = y_1, \dots, y_n$.

СУБЪЕКТИВНОЕ МОДЕЛИРОВАНИЕ ДАННЫХ ИЗМЕРЕНИЙ И КРИТЕРИИ ОПТИМАЛЬНОСТИ СУБЪЕКТИВНЫХ РЕШЕНИЙ

Сглаживающий сплайн $\sigma(\cdot) \in S_\rho^q$ позволяет представить результаты измерений $y_i, i = 1, \dots, n$, в (3) в виде двух слагаемых: первое – «гладкое» $\sigma(t_i), i = 1, \dots, n$, оценивающее зависимость температуры воды от времени, второе – «негладкое» $y_i - \sigma(t_i), i = 1, \dots, n$, моделирующее зависимость от времени шума.

Субъективные правила принятия решений состоят в таком выборе двух параметров ρ и q , при котором выполнены два критерия: функция $\sigma(t_i)$ при всех $t_i, i = 1, \dots, n$, – гладкая, без следов шума, свойственных разностям $\mu_i, i = 1, \dots, n$, а разности $\mu_i, i = 1, \dots, n$, не содержат фрагментов гладкой зависимости, свойственных функции $\sigma(t_i), i = 1, \dots, n$, и «выглядят» как шум. При этом критерий точности решений задач субъективного моделирования температуры и интерпретации данных измерений априори не используется.

СУБЪЕКТИВНОЕ МОДЕЛИРОВАНИЕ ДАННЫХ ИЗМЕРЕНИЙ И КРИТЕРИИ ОПТИМАЛЬНОСТИ СУБЪЕКТИВНЫХ РЕШЕНИЙ

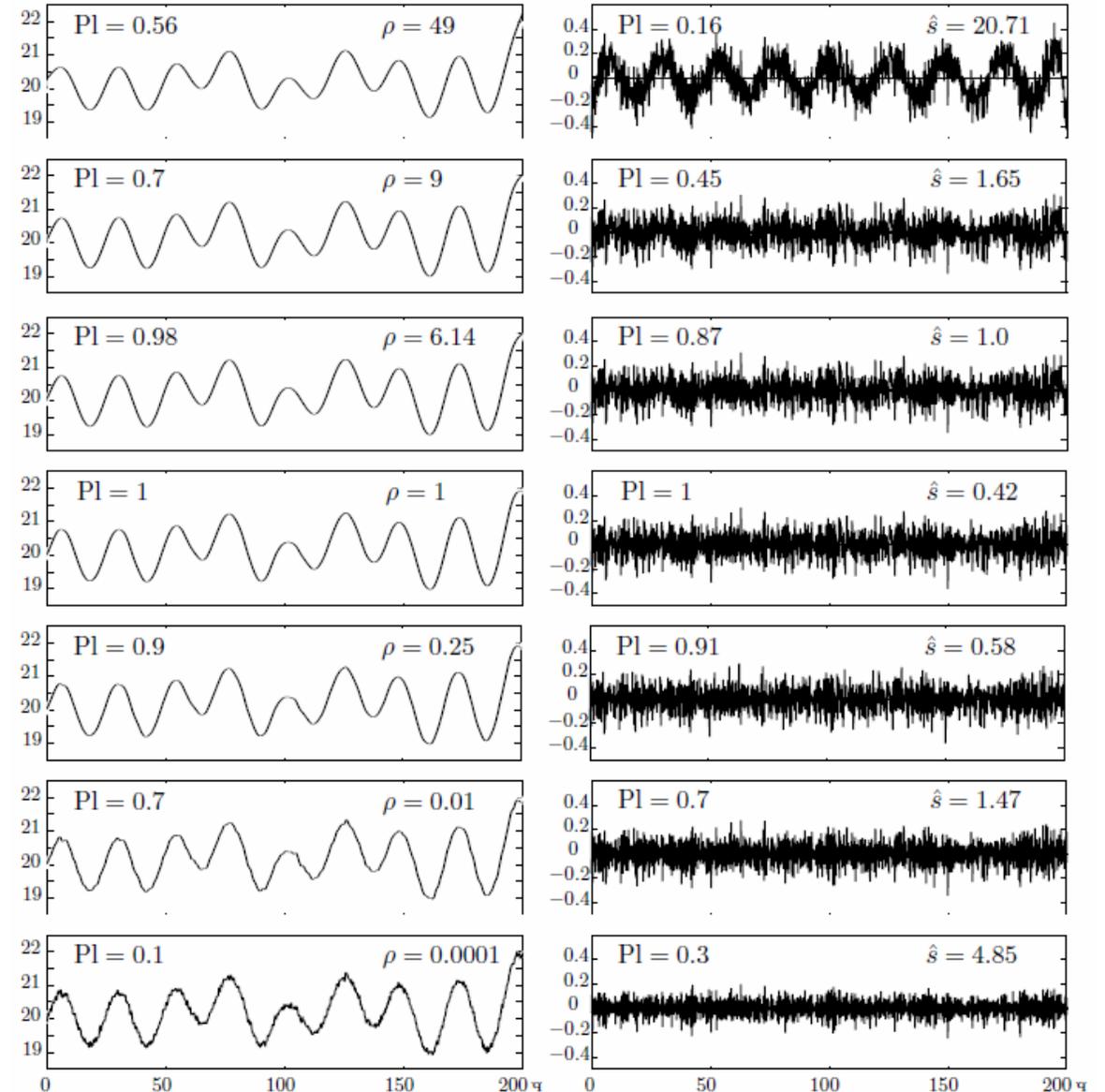
Поскольку «гладкость» – естественно воспринимаемое субъективно качество графика функции, присутствующее в обоих критериях, выберем ее предельное значение, определенное величиной q , и зафиксируем, а подбирать будем фактор ρ , одновременно рассматривая и сравнивая для каждого значения ρ графики зависимостей $\sigma(t_i)$, $i = 1, \dots, n$, и μ_i , $i = 1, \dots, n$, добиваясь выполнения сформулированных критериев для обоих графиков. После этого тем графикам сплайна и разностей, для которых выполнены оба критерия, присвоим значения их правдоподобий, равные единице, а остальным графикам присвоим правдоподобия, субъективно характеризующие их истинности как графиков зависимостей от времени температуры воды и шума. Эти же значения правдоподобий присвоим и значениям сглаживающего фактора ρ , определившим графики сплайна $\sigma(\cdot)$ и разностей μ . как оценок температуры воды и шума (см. рис. 2). Заметим, что равные единице значения правдоподобий сплайна и разностей при правильном выборе значения q должны иметь равные значения ρ !

СУБЪЕКТИВНОЕ МОДЕЛИРОВАНИЕ ДАННЫХ ИЗМЕРЕНИЙ И КРИТЕРИИ ОПТИМАЛЬНОСТИ СУБЪЕКТИВНЫХ РЕШЕНИЙ

В верхней строке левый график может моделировать зависимость от времени, но верхний правый график промодулирован графиком слева и не может моделировать график шума.

В последней строке левый график не может служить графиком температуры, поскольку на нём видны следы шума, но правый график мог бы моделировать шум.

В верхнем правом углу графиков разностей приведены реальные значения с. к. погрешностей $\hat{s}_\alpha = \|f(\cdot) - \sigma^\alpha(\cdot)\|^2$, $\alpha = 1, \dots, 7$. Видно, что минимум значений с.к. погрешностей достигается на оптимальных субъективных интерпретациях максимального правдоподобия. Этот факт свидетельствует об адекватности полученных субъективных моделей измерения и интерпретации данных измерений, поскольку при построении моделей максимизация точности интерпретации априори не использовалась, а обнаружена экспериментально при использовании субъективных моделей.



ДААННЫЕ С ПРОПУСКАМИ

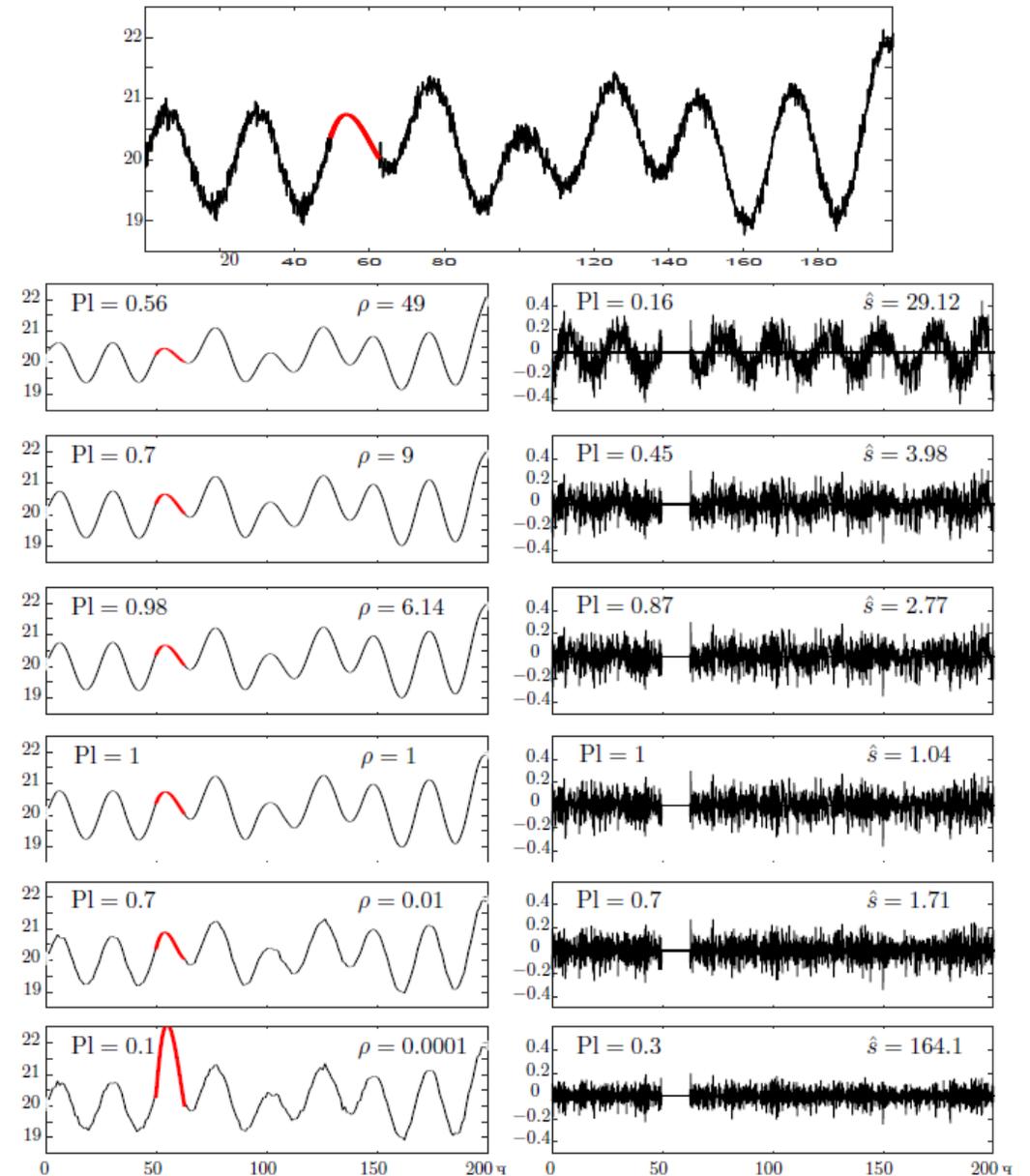
В верхней строке левый график может моделировать зависимость от времени, но верхний правый график промодулирован графиком слева и не может моделировать график шума.

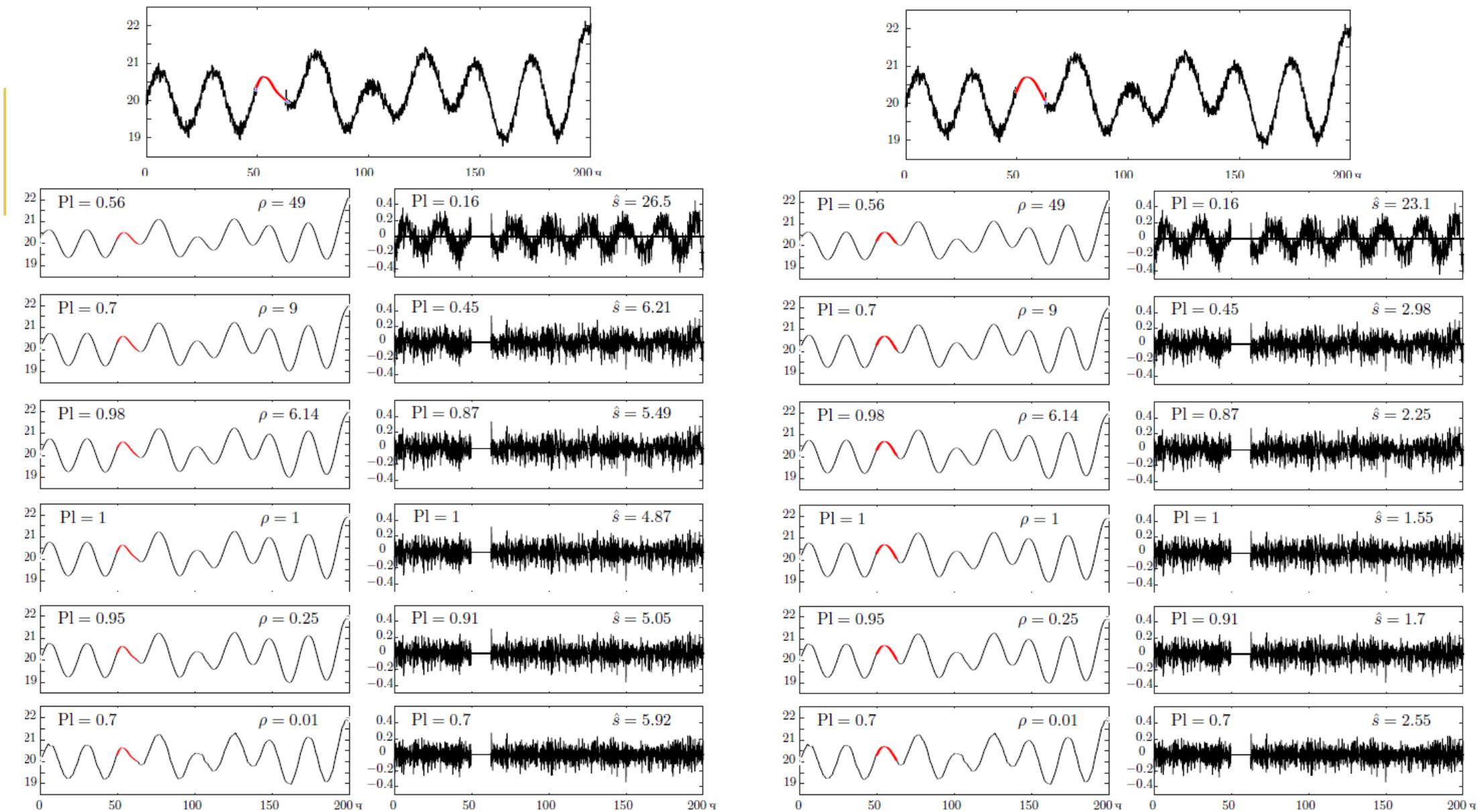
В последней строке левый график не может служить графиком температуры, поскольку на нём видны следы шума, но правый график мог бы моделировать шум.

Пропуск заполнен значениями сплайна, определенного значением $\rho_{opt} = 1$ сглаживающего фактора.

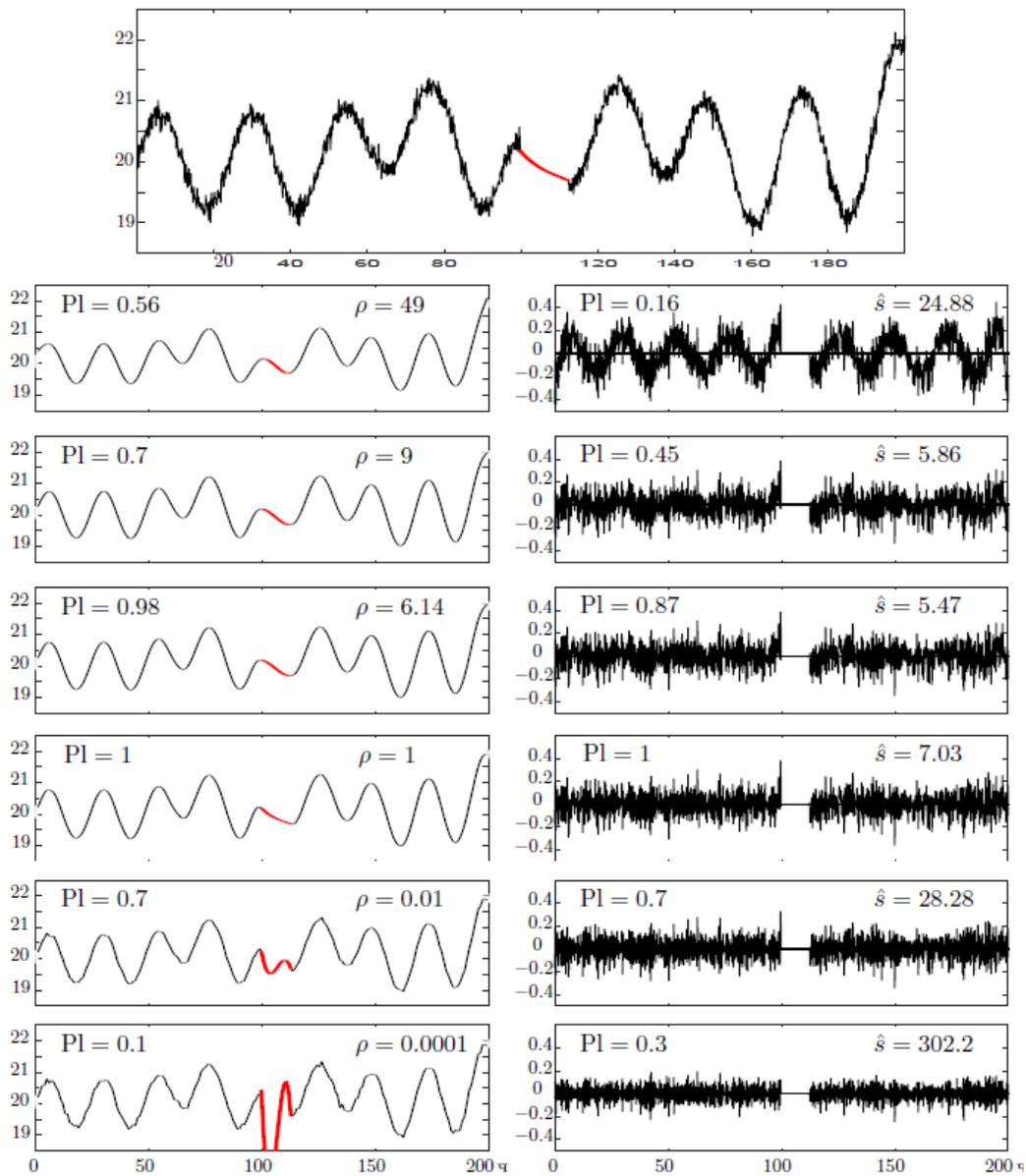
В верхнем правом углу графиков разностей приведены реальные значения с. к. погрешностей

$\hat{s}_\alpha = ||f(\cdot) - \sigma^\alpha(\cdot)||^2$, $\alpha = 1, \dots, 7$. Видно, что минимум значений с.к. погрешностей достигается на оптимальных субъективных интерпретациях максимального правдоподобия. Этот факт свидетельствует об адекватности полученных субъективных моделей измерения и интерпретации данных измерений, поскольку при построении моделей максимизация точности интерпретации априори не использовалась, а обнаружена экспериментально при использовании субъективных моделей.

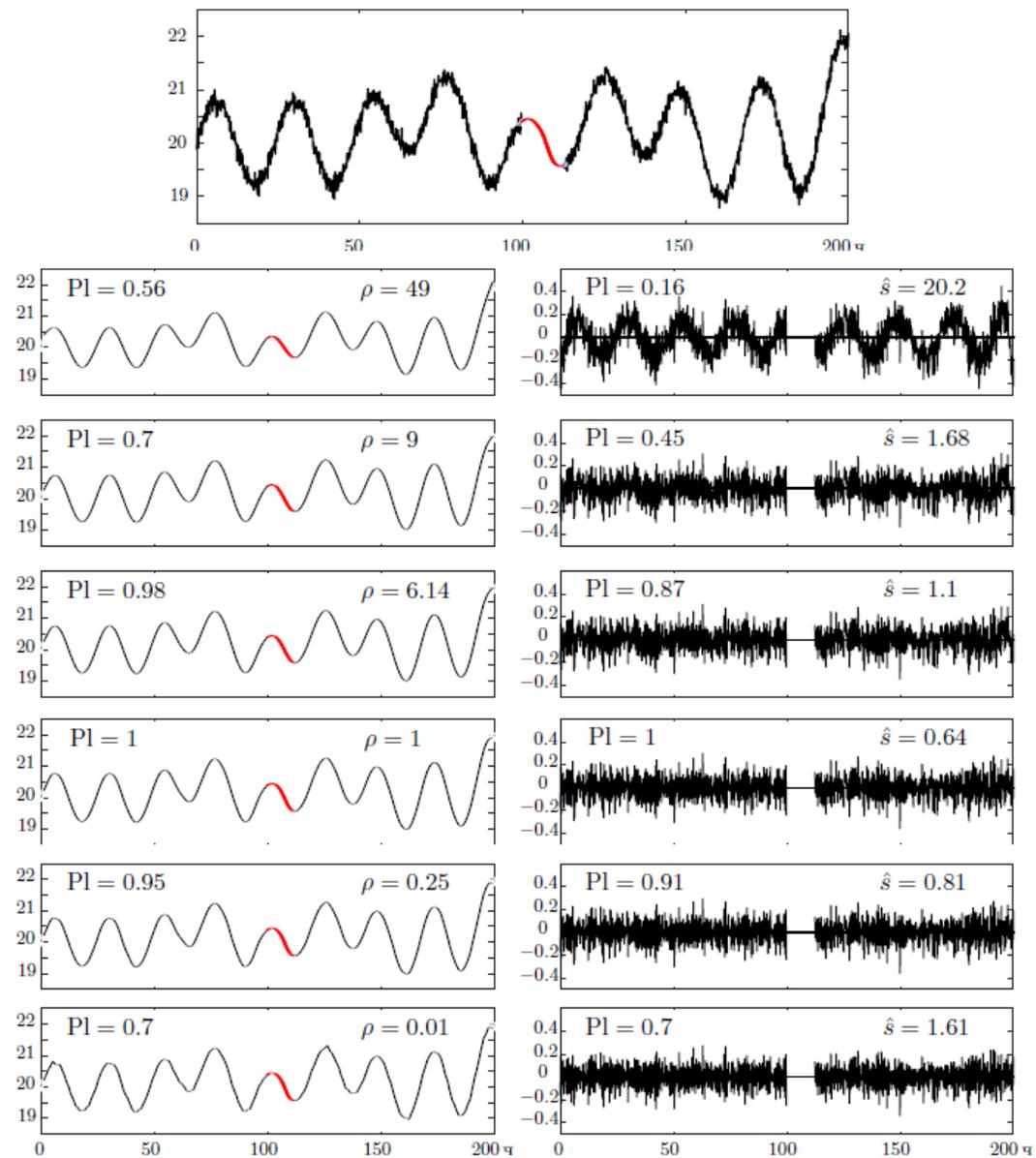




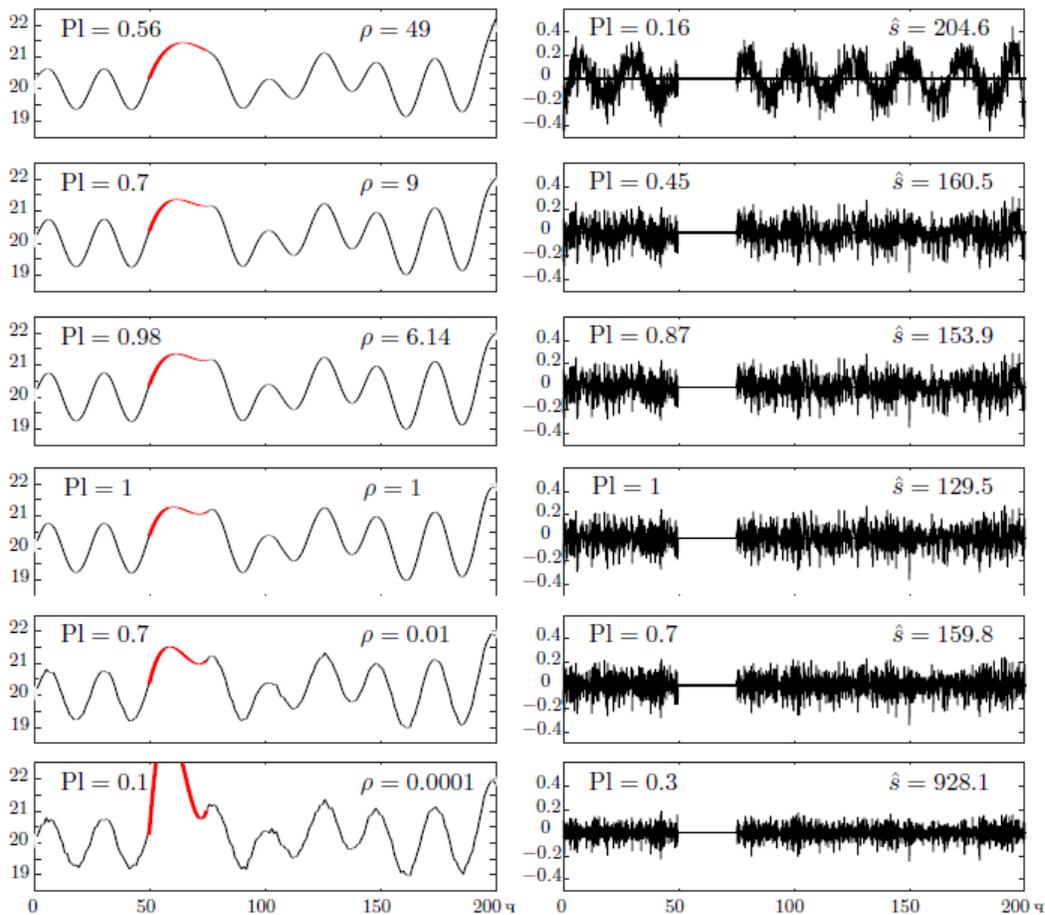
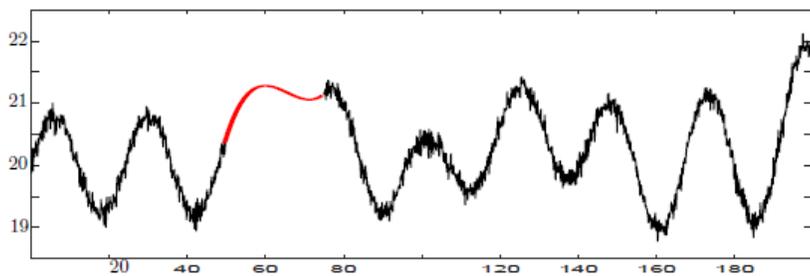
Пропуск шириной 100 точек заполнен субъективно модельером-исследователем.



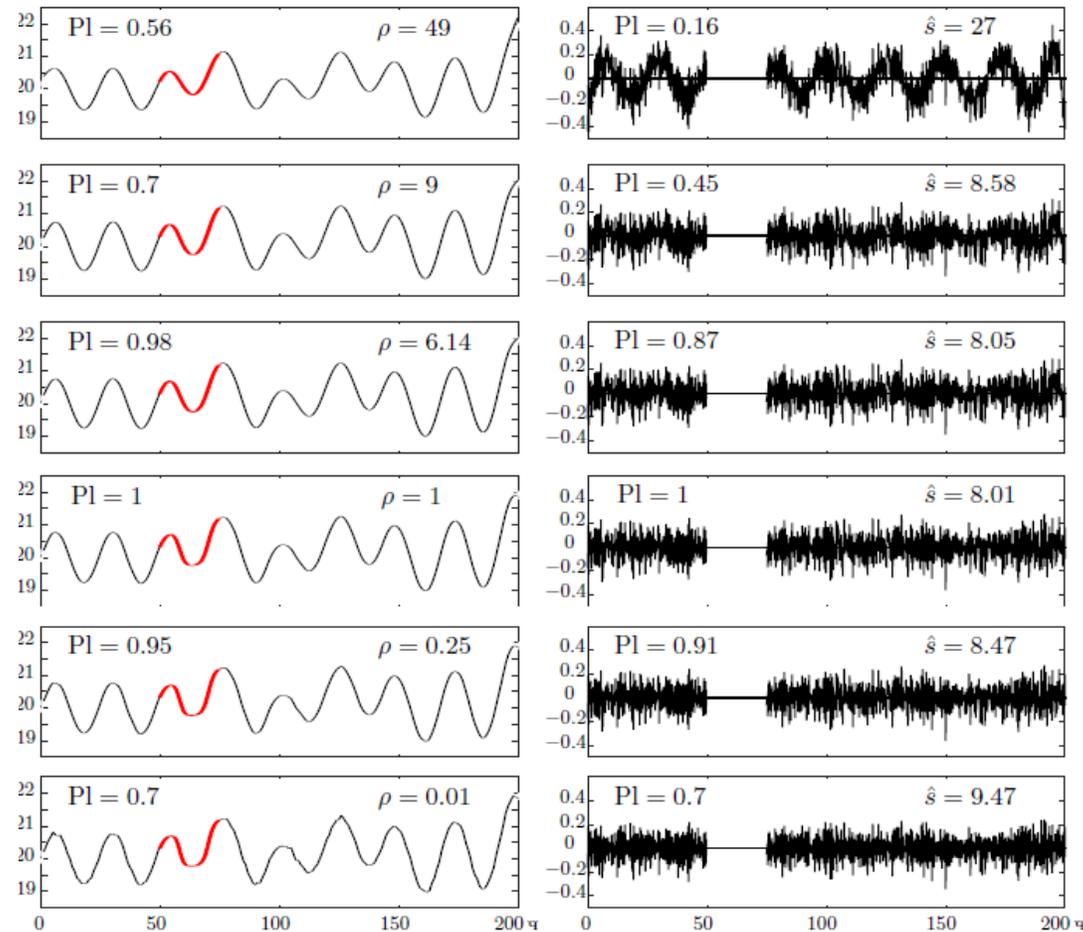
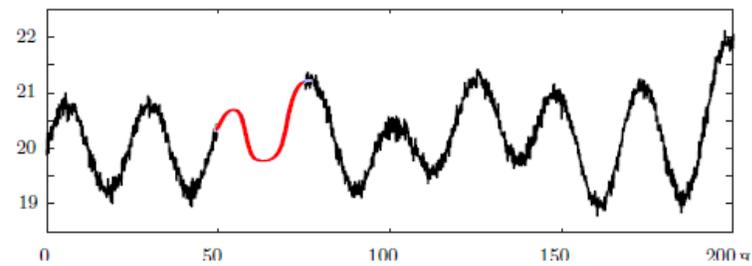
Пропуск шириной 100 точек заполнен значениями сплайна, определенного значением $\rho_{\text{opt}} = 1$ сглаживающего фактора.



Пропуск шириной 100 точек заполнен субъективно модельером-исследователем.



Пропуск шириной 200 точек заполнен значениями сплайна, определенного значением $\rho_{opt} = 1$ сглаживающего фактора.



Пропуск шириной 200 точек заполнен субъективно модельером-исследователем с учетом его знаний о суточных колебаниях температуры.

1. *Pyt'ev Y. P.* Modeling of subjective judgments made by a researcher-modeler about the model of the research object // *Mathematical Models and Computer Simulations*. — 2013. — Vol. 5, no. 6. — P. 538–557.

2. *Ю. П. Пытьев.* Вероятность, возможность и субъективное моделирование в научных исследованиях. Математические и эмпирические основы, приложения. Москва. ФИЗМАТЛИТ, 2018, 268 стр.

3. *Ю. П. Пытьев, О. В. Фаломкина, С. А. Шишкин, А. И. Чуличков.* Математический формализм субъективного моделирования // *Машинное обучение и анализ данных*, Москва, 2018. — Т. 4, № 2. — С. 108–121.

4. *Yu. P. Pyt'ev, O. V. Falomkina, S. A. Shishkin.* Subjective Restoration of Mathematical Models for a Research Object, Its Measurements, and Measurement-Data Interpretation // *Pattern Recognition and Image Analysis*, № 4, 2019 (принята к печати).