Применение сэмплирования из условного распределения для оценки предсказательной способности квалифицированных нелинейных моделей со смешанными эффектами

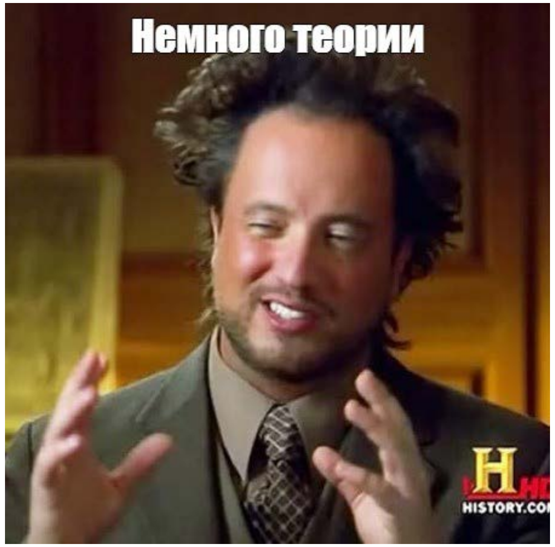
Анна Мишина, Кирилл Жуденков

Ноябрь, 2023

XV конференция «Математические модели и численные методы в биологии и медицине»

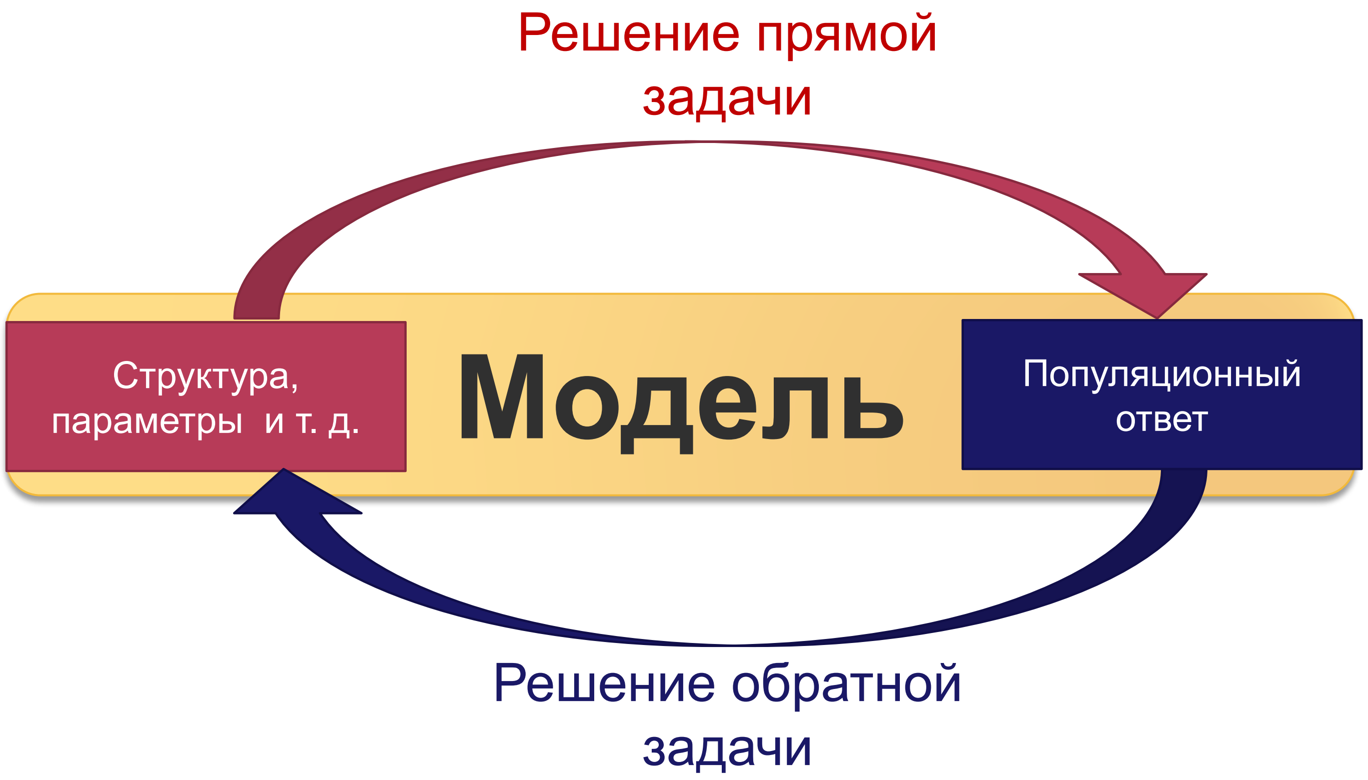
Структура доклада

1. Актуальность
2. Постановка задачи по моделированию
3. Алгоритм оптимизации параметров SAEM
4. Алгоритм оптимизации индивидуальных параметров
5. Процедура сэмплирования из условного распределения в ПО Monolix Suite
6. Процедура валидации
7. План исследования
8. Результаты полученные на данных клинических испытаний
9. Выводы

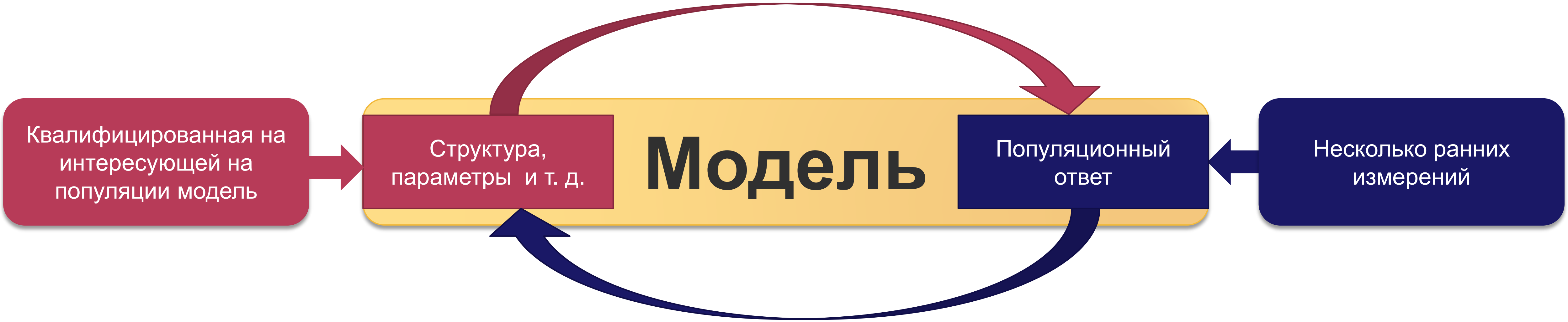


IM&S

Актуальность



Актуальность



Цель работы:
Подбор оптимального алгоритма построения долговременных предсказаний для измеряемых во времени биомаркеров на основе данных промежуточных измерений и квалифицированной модели

IM&S

Постановка задачи в вероятностном виде

$1 \; individual, n \; observations$

$$\vec{y} = (y_j, 1 \le j \le n)$$

$$\vec{t} = (t_j, 1 \le j \le n)$$

$$y = f(t; \vec{\varphi}), \qquad where$$

$$f(t; \vec{\varphi}) = f(t; Base, k_g, k_d, \lambda) = Base \cdot e^{k_g \cdot t \cdot \frac{k_d}{\lambda} \cdot (1 - e^{-\lambda \cdot t})}$$

$$y_j = f(t_j; \vec{\varphi}) + e_j, 1 \le j \le n$$

$Assume \; that \; residual \; errors \; (e_j) \; are \; independent \; and \; normally \; distributed \; with \; constant \; variance \; a^2$:

$$y_j \sim \mathcal{N}(f(t_j; \vec{\varphi}), a^2)$$

$$p_{\vec{y}} = p(\vec{y}; \vec{\varphi}, \vec{t}) = \prod_{j=1}^{n} p_{y_j}(y_j; \vec{\varphi}, t_j)$$

Постановка задачи в вероятностном виде

$\mathrm{N}\ individuals$

$$\boldsymbol{y} = \left(y_{ij}, 1 \le i \le N, 1 \le j \le n_j\right)$$
$$\boldsymbol{t} = (t_{ij}, 1 \le i \le N, 1 \le j \le n_j)$$
$$\boldsymbol{\varphi} = (\overrightarrow{\varphi_i}, 1 \le i \le N)$$

$$p_{\boldsymbol{\varphi}} = (\cdot;\ \vec{\theta})$$
$$p\left(\boldsymbol{y}, \boldsymbol{\varphi};\ \vec{\theta}, \boldsymbol{t}\right) = p(\boldsymbol{y}|\boldsymbol{\varphi}; \boldsymbol{t}) \cdot p(\boldsymbol{\varphi}; \vec{\theta})$$

$$y_{ij} \sim \mathcal{N}\left(f\left(t_{ij};\ \overrightarrow{\varphi_i}\right), {a_i}^2\right)$$

$Assume\ the\ same\ residual\ error\ for\ all\ patients\ and\ lognormal\ distributions\ for\ Base, k_g, k_d\ and\ \lambda$:

$$a_i = a$$
$$\log(Base_i) \sim \mathcal{N}\left(\log\left(Base_{pop}\right), \omega_{Base}^2\right)$$
$$\log\left(k_{gi}\right) \sim \mathcal{N}\left(\log\left(k_{g,pop}\right), \omega_{k_g}^2\right)$$
$$\ldots$$

Алгоритм оптимизации параметров SAEM

$By\ definition, the\ maximum\ likelihood\ estimator\ of\ \theta\ maximizes$:

$$\mathcal{L}_{\boldsymbol{y}}(\vec{\theta}) = p(\boldsymbol{y}; \vec{\theta}) = \int p(\boldsymbol{y}, \boldsymbol{\varphi}; \vec{\theta}) d\boldsymbol{\varphi}$$

EM algorithm

E-step: Evaluate the quantity

$$Q_k^{EM}(\vec{\theta}) = \mathbb{E}(\log p(\boldsymbol{y}, \boldsymbol{\varphi}; \vec{\theta}) | \boldsymbol{y}; Q_{k-1}^{EM})$$

M-step: Update the estimation of $\vec{\theta}$

$$\vec{\theta}_k^{EM} = \arg\max_{\vec{\theta}} Q_k^{EM}(\vec{\theta})$$

SAEM algorithm

Simulation step: For $i = 1,2, \dots, N$, draw $\vec{\varphi}_i^{(k)}$ from the conditional distribution $p(\vec{\varphi_i} | \vec{y_i}; \overrightarrow{\theta_{k-1}})$

Stochastic approximation: Update $Q_{k-1}(\vec{\theta})$ according to

$$Q_k(\vec{\theta}) = Q_{k-1}(\vec{\theta}) + \gamma_k (\log p(\boldsymbol{y}, \boldsymbol{\varphi}^{(k)}; \vec{\theta}) - Q_{k-1}(\vec{\theta})),$$

where (γ_k) is a decreasing sequence of positive numbers such that $\gamma_1 = 1$

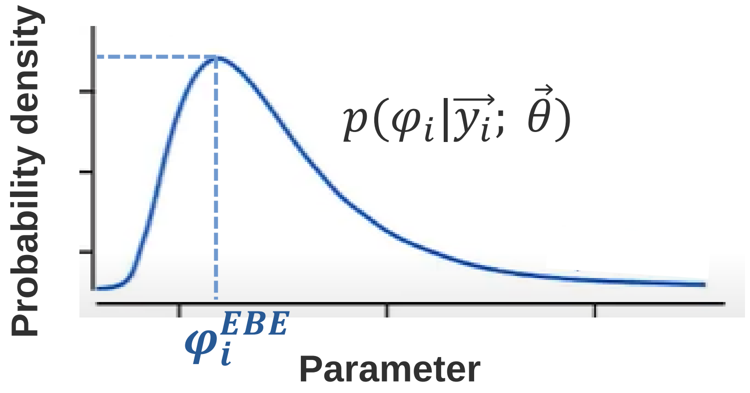
Maximization step: Update $\overrightarrow{\theta_k}$ according to

$$\vec{\theta}_k = \arg\max_{\vec{\theta}} Q_k(\vec{\theta})$$

Lavielle, M. Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools.

Monolix: оценка индивидуальных параметров (EBEs)

$$EBE = mode\ of\ conditional\ distribution$$



$$\varphi_i^{mode} = \underset{\varphi_i}{\arg\max}\, p(\varphi_i|\overrightarrow{y_i};\vec{\theta})$$

$$p(\varphi_i|\overrightarrow{y_i};\vec{\theta}) = \frac{p(\overrightarrow{y_i}|\varphi_i)\cdot p(\varphi_i)}{p(\overrightarrow{y_i})}$$

$$\varphi_i^{mode} = \underset{\varphi_i}{\arg\max}\, (p(\overrightarrow{y_i}|\varphi_i)\cdot p(\varphi_i))$$

Maximization done via Nelder-Mead simplex algorithm

https://monolix.lixoft.com/tasks/ebes/

Monolix: сэмплирование из условного распределения

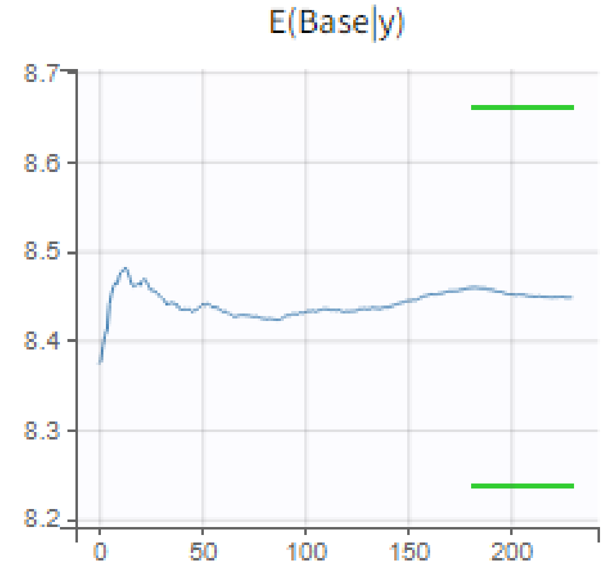
With Monolix we could also generate values for the individual parameters from conditional distribution

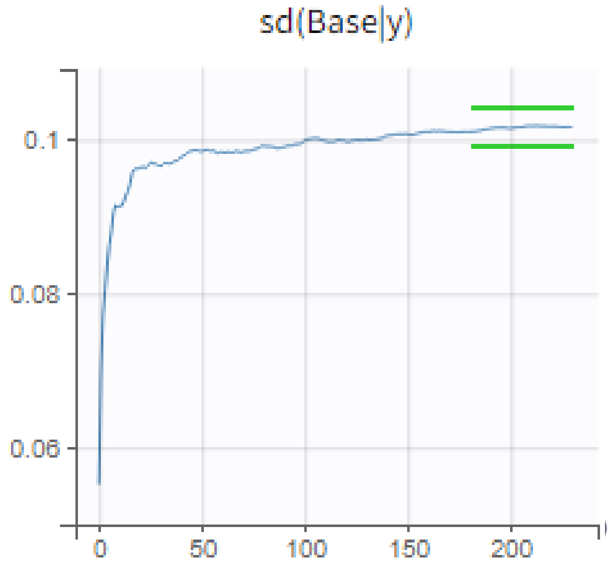
$$p\left(\varphi_i \middle| \vec{y_i}; \vec{\theta}\right) = \frac{p(\vec{y_i}|\varphi_i) \cdot p(\varphi_i)}{p(\vec{y_i})}$$

The Metropolis-Hastings algorithm is used to simulate a sequence of individual parameters $\left(\varphi_i^{(l)}, l = 1,2, \dots\right)$

The simulated sequence can then be used for empirically estimating the conditional distribution $p_{\varphi_i|y_i}$ *and the conditional means* $\mathbb{E}\left(\varphi_i \middle| \vec{y_i}; \vec{\theta}\right)$ *and conditional standard deviations* $\mathrm{sd}(\varphi_i|\vec{y_i}; \vec{\theta})$

The sampling stops when these sequences remain in an interval of a given amplitude for certain number of iterations





Lavielle, M. Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools.

Процедура валидации

Загрузите проект с квалифицированной на тренировочных данных моделью

Используете в качестве начальных значения идентифицированные параметры

Зафиксируйте их

Загрузите часть валидационных данных с промежуточными измерениями

Сохраните новый проект

Запустите SAEM и EBEs функции

- 1 набор параметров на каждого пациента, оцененный как мода условного распределения

- 1 набор параметров, оцененный в ходе работы SAEM алгоритма

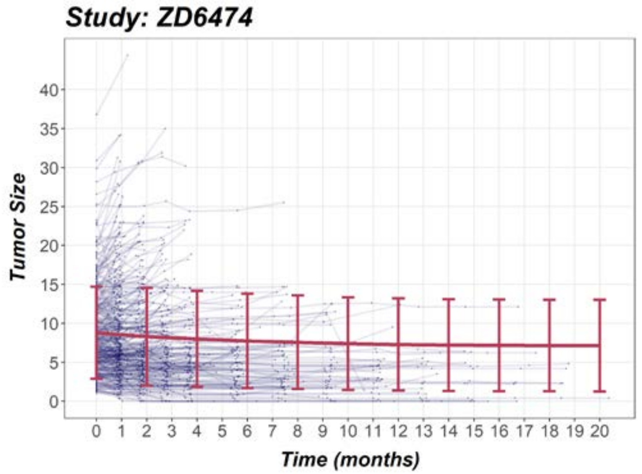
Запустите SAEM и Conditional Distribution функции

- n наборов параметров на каждого пациента, где n – регулируемый параметр (у нас n=50)

План исследования

Данные

381 пациент
1691 измерение



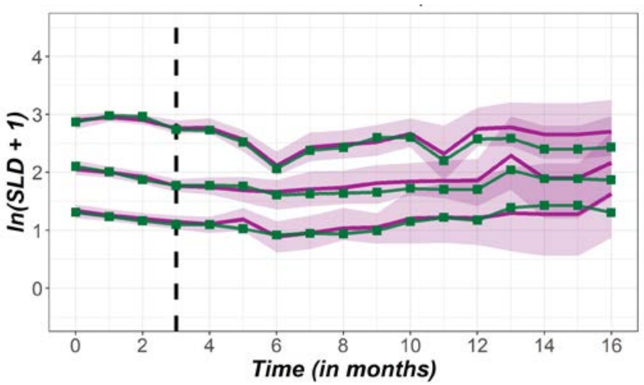
Разбиение данных (250/131)

Квалификация

Parameter	Value	R.S.E. (%)
Base_pop	7.24	4.33
k_d_pop	0.043	36.7
k_g_pop	0.033	16.7
la_pop	0.69	27.0
omg_Base	0.68	4.58
omg_k_d	1.89	13.4
omg_k_g	1.4	7.95
omg_la	1.44	16.0
corr_kd_Base	-0.36	23.2
a	0.1	3.03

Валидация

id	Base	k_d	k_g	la
1	6.74	0.021	0.007	1.3
2	…	…	…	…



50x

IM&S

Результаты

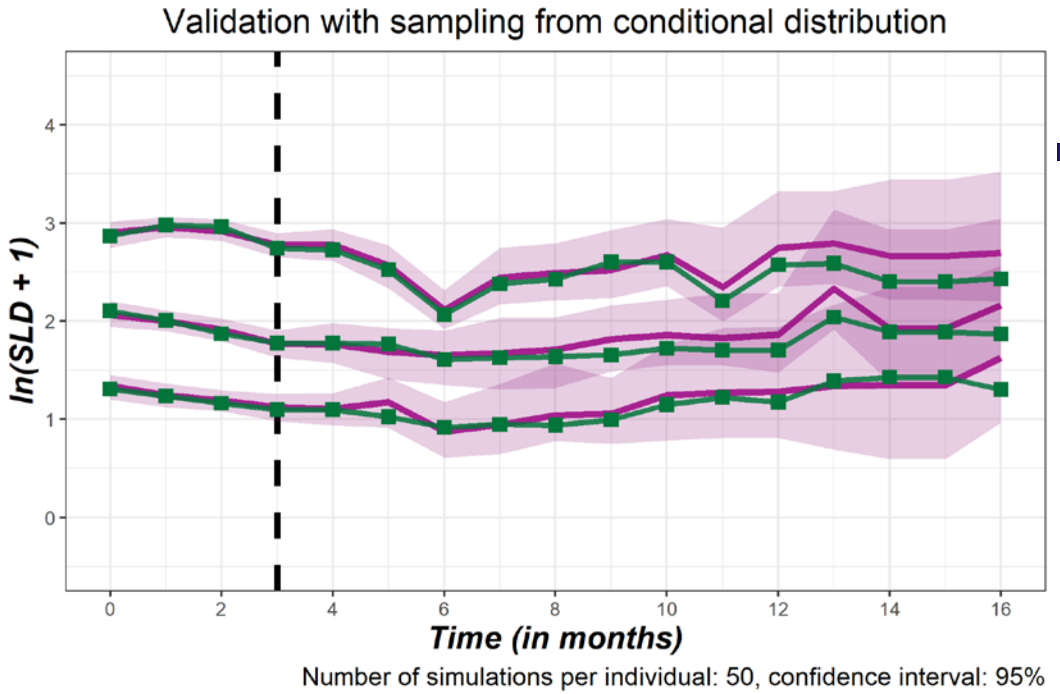
- Предложенный подход позволяет построить доверительный интервал

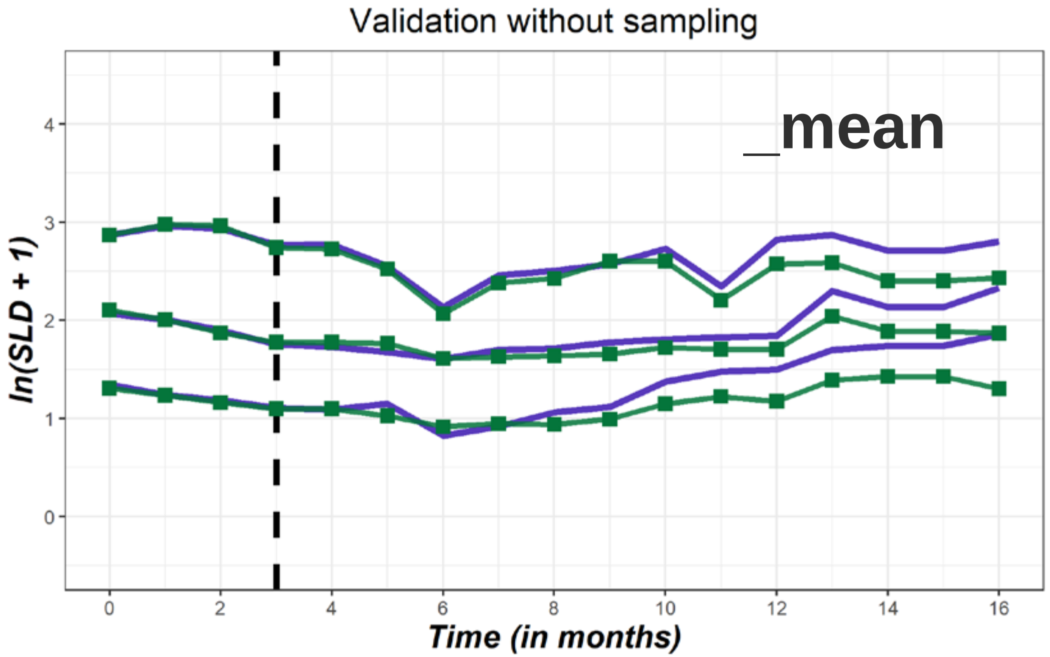
Экспериментальные значения SLD

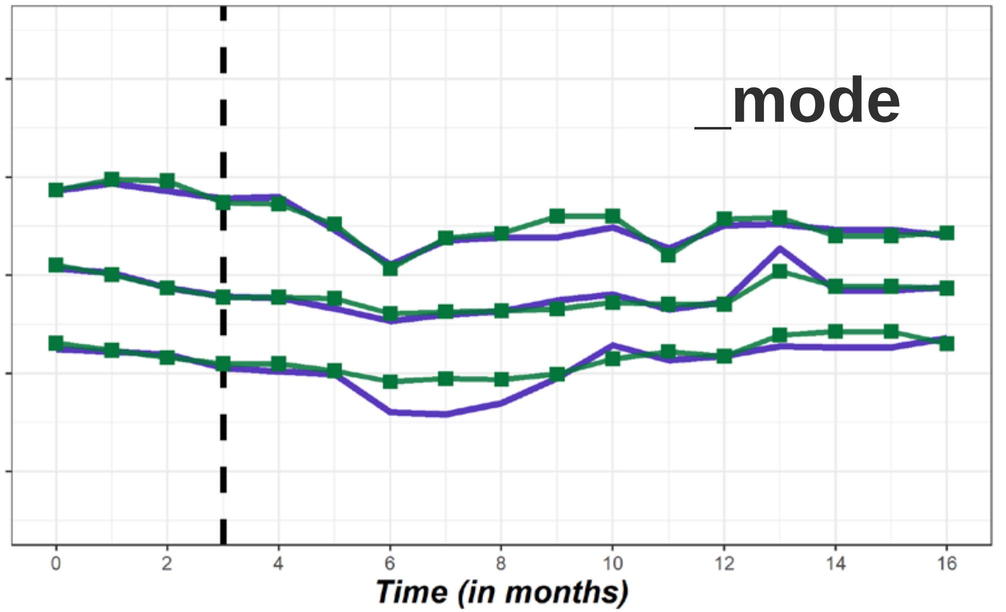
Предсказанный SLD профиль без применения сэмплирования

Предсказанный SLD профиль с применением сэмплирования

95% доверительный интервал вокруг предсказания, определенный с применением сэмплирования





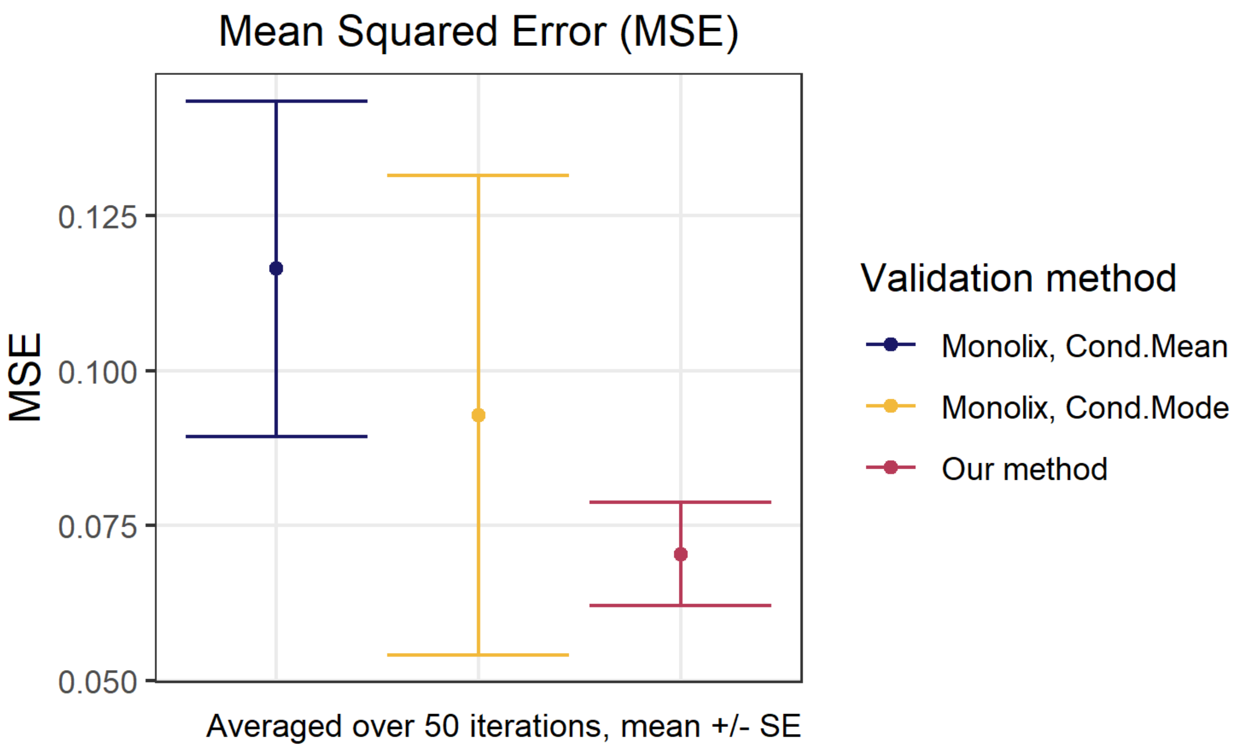


Результаты

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Предложенная методология обеспечивает прогнозы, не уступающие по точности тем, что получаются при использовании EBEs

- При этом точность предложенной методологии оказывается значимо выше, чем в подходе с использованием оценки условного среднего

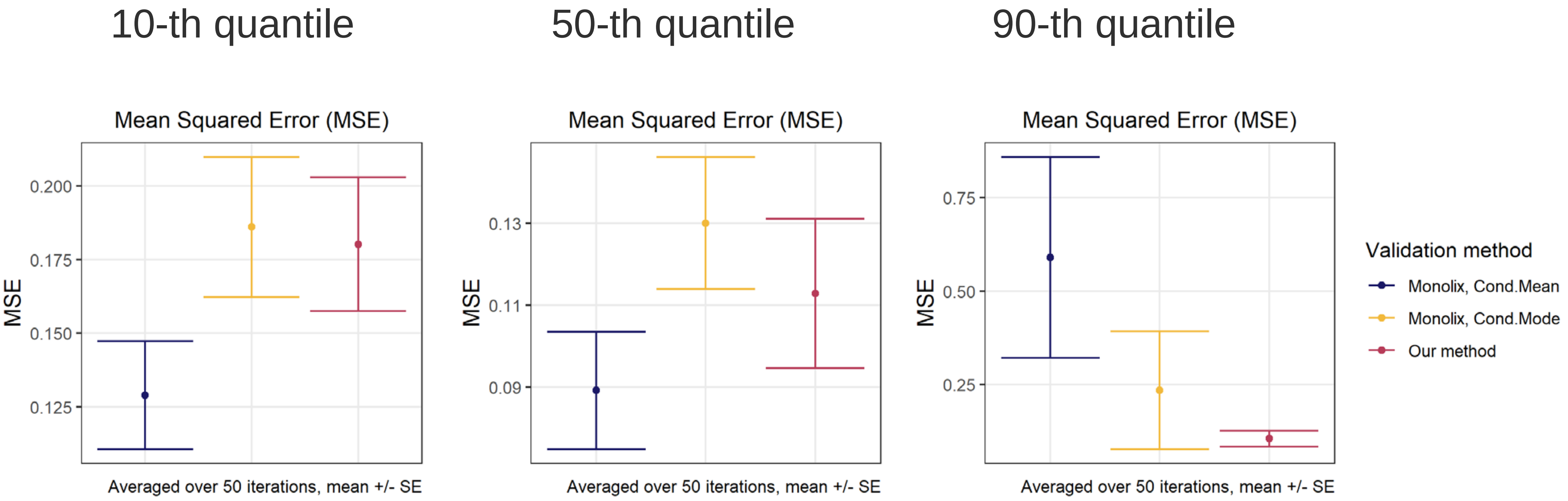


Выводы

1. Предложенная методология позволяет оценивать доверительные интервалы вокруг предсказаний
2. Среднее среднеквадратичной ошибки для предсказаний в случае адаптированной методологии оказывается ниже, чем у рекомендованной Мононикс

BACK UP

Разные метрики с разной точностью предсказывают разные квантили



IM&S