

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук  
Образовательная программа «Прикладная математика и информатика»

УДК 004.032.26

**Отчет об исследовательском проекте на тему:**

**Прогноз временных рядов с помощью диффузионных моделей**

**Выполнил студент:**

группы БПМИ213, 3 курса                    Филатов Егор Дмитриевич

**Принял руководитель проекта:**

Гущин Михаил Иванович  
Старший научный сотрудник  
Факультет компьютерных наук НИУ ВШЭ

Москва 2024

# Содержание

<b>Аннотация</b>	<b>3</b>
<b>1 Введение</b>	<b>4</b>
1.1 Описание предметной области . . . . .	4
1.2 Постановка задачи . . . . .	4
<b>2 Обзор и сравнительный анализ источников</b>	<b>5</b>
2.1 Другие методы заполнения временных рядов . . . . .	5
2.2 CSDI . . . . .	5
2.3 GP-VAE . . . . .	6
2.4 SAITS . . . . .	7
2.5 BRITS . . . . .	7
2.6 Другие сравнительные исследования . . . . .	8
<b>3 Постановка задачи заполнения временных рядов</b>	<b>8</b>
<b>4 Условия экспериментов</b>	<b>9</b>
<b>5 Результаты</b>	<b>11</b>
<b>6 Выводы</b>	<b>13</b>
<b>Список литературы</b>	<b>15</b>
<b>А Примеры графиков предсказаний</b>	<b>17</b>

## **Аннотация**

Данная курсовая работа посвящена исследованию заполнения пропусков во временных рядах с использованием диффузионных моделей. В работе в качестве основного рассматривается метод CSDI (Conditional Score-based Diffusion model for Imputation)[15], который представляет собой эффективный подход к прогнозированию временных рядов и заполнению пропусков в них на основе диффузионных моделей. Основной целью работы является изучение применимости метода CSDI для заполнения пропусков в реальных временных рядах различной природы и сравнение его эффективности с другими стандартными методами прогнозирования, такими как KNN, и нейросетевыми моделями, такими как BRITS[1], SAITS[4], GP-VAE[5]. В качестве основных методов работы используются анализ литературы по теме, поиск и обработка данных временных рядов, программная реализация экспериментов с моделями прогнозирования, а также статистический анализ результатов. Полученные результаты и выводы будут иметь значение для применения диффузионных моделей в прогнозировании временных рядов и дальнейших исследований в этой области.

## **Ключевые слова**

Временные ряды, прогнозирование временных рядов, заполнение пропусков во временных рядах, диффузионные модели, CSDI, машинное обучение, сравнительный анализ

# **1 Введение**

## **1.1 Описание предметной области**

В современном мире прогнозирование временных рядов играет ключевую роль в различных областях, включая экономику, финансы, климатологию, здравоохранение, и многие другие. Эффективные модели прогнозирования позволяют предсказывать будущие значения временных рядов на основе прошлых наблюдений, что является критически важным для принятия обоснованных решений. Однако, существующие методы прогнозирования временных рядов все еще сталкиваются с вызовами, такими как неустойчивость, недостаточная точность и неспособность обрабатывать пропуски данных. С появлением диффузионных моделей в области прогнозирования временных рядов открылись новые перспективы в предсказании сложных динамических систем. В данной статье проводится сравнительный анализ моделей прогнозирования временных рядов. Особое внимание уделяется модели CSDI (Conditional Score-based Diffusion model for Imputation), которая предлагает инновационный подход к заполнению пропусков в данных. Проблема прогнозирования временных рядов с пропусками данных имеет огромное практическое значение, поскольку реальные наборы данных часто содержат недостающие значения из-за различных причин, таких как ошибки измерения или отсутствие данных.

## **1.2 Постановка задачи**

Поставленные задачи включают в себя анализ эффективности метода CSDI в прогнозировании временных рядов различной природы, сравнение его с другими методами прогнозирования и оценку достоверности полученных прогнозов.

Целью данной работы является исследование эффективности модели CSDI в сравнении с другими методами заполнения пропусков и моделями прогнозирования временных рядов. Актуальность данного исследования обусловлена необходимостью разработки более точных и надежных методов прогнозирования временных рядов, способных учитывать сложные динамические взаимосвязи в данных. Результаты этого исследования могут иметь значительное влияние для различных областей, где прогнозирование временных рядов играет ключевую роль, помогая улучшить точность и надежность прогнозов, а также для дальнейших исследований в области диффузионного моделирования временных рядов.

## 2 Обзор и сравнительный анализ источников

### 2.1 Другие методы заполнения временных рядов

В свете важности задачи прогнозирования временных рядов и заполнения пропусков в них во многих прикладных задачах, для её решения существует множество реализованных решений, пройдемся по основным из них. Одними из первых методов в силу своей простоты были статистические методы, такие как KNN (k-nearest neighbours)[14] или линейная интерполяция, которые но они имели низкое качество и неспособны выделять сложные нелинейные зависимости во временных рядах. На смену им пришли модели машинного обучения, такие как линейная регрессия[9] и решающие деревья. Эти методы пришли с появлением машинного обучения, они уже оказались способны делать лучшие предсказания, но их точности все равно было недостаточно для большинства прикладных задач. Лучше всего себя показали модели глубинного обучения, как RNN (recurrent neural network)[1], Transformer-based архитектуры[4] и VAE (variational autoencoder)[5]. Они оказались отлично применимы к временным рядам, поскольку улавливали последовательную структуру рядов и могли находить сложные закономерности в силу сложности своей архитектуры и лучшей способности к генерации данных.

Но исследования на этом не остановились, и с ростом популярности диффузионных моделей в последние годы адаптация их для временных рядов была только вопросом времени, так как они проявляют выдающуюся производительность при генерации данных, сильно похожих на реальные наблюдаемые данные.

### 2.2 CSDI

В качестве исследуемой модели была выбрана модель CSDI, так как эта модель является первой диффузионной сетью, специализирующейся на заполнении пропусков во временных рядах, пришедшей на смену рекуррентным моделям. По сути CSDI — это DDPM (Denoising Diffusion Probabilistic Model)[8], то есть в процессе обучения модель итеративно зашумляет известные данные с помощью нормально распределенного шума, чтобы свести их к нормальному распределению, при этом она учится восстанавливать зашумление, а в процессе генерации пропущенных значений восстанавливает их из гауссовского шума, проводя обратный к зашумлению процесс, таким образом модель приближает распределение реальных данных и может иметь высокую точность при генерации. Так как наша задача состоит в том, чтобы заполнить пропуски по имеющимся данным, а не просто сгенериро-

вать какой-то временной ряд, в модель добавляется обуславливание на имеющиеся данные и модель учится приближать шум с этим условием, минимизируя функцию потерь

$$\mathcal{L}(\theta) = \mathbb{E}_{x_0 \sim q(x_0), \epsilon \sim \mathcal{N}(0, I), t} \|\epsilon - \epsilon_\theta(x_t^{ta}, t \mid x_0^{co})\|_2^2, \quad (1)$$

где  $x_t^{ta}$  — зашумленные тестовые данные, которые нам нужно восстановить,  $x_0^{co}$  — обуславливающие известные данные,  $\epsilon_\theta$  — предсказания нашей нейросети.

Архитектуру, а именно несколько последовательных блоков сверток, соединенных с выходом через skip-connections, CSDI наследует от DiffWave[10], за исключением того, что некоторые свертки в CSDI заменены на Transformer attention-слои, так как по результатам исследования они показали себя лучше.

Большим плюсом её является то, что она обучается она в self-supervised режиме, в датасете с пропущенными значениями она маскирует часть известных данных и восстанавливает зависимости по известным. Описание, метод работы и параметры модели взяты из статьи[15], где и была предложена данная модель.

## 2.3 GP-VAE

Стоит рассказать подробнее об остальных моделях, которые будут рассматриваться в этой работе. GP-VAE (Gaussian Process Variational Autoencoder)[5] использует идею оригинального VAE[2], модель пытается спроектировать временной ряд на пространство меньшей размерности, в которой можно было бы смоделировать гауссовский процесс с помощью которого можно будет предсказывать отсутствующие переменные во временном ряду. Сжатие временного ряда происходит с помощью сверточной сети, а генерация переменных из гауссовского процесса с помощью полносвязной сети. Зарисовку архитектуры можно увидеть на Рисунке 2.1

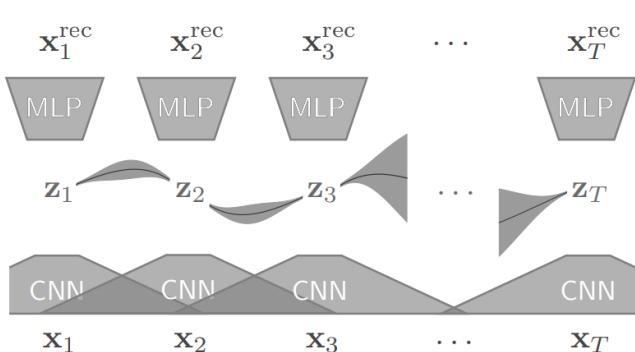


Рис. 2.1: Архитектура GP-VAE

## 2.4 SAITS

SAITS (Self-Attention-Based Imputation for Time Series)[4] — это модель, основанная на архитектуре self-attention Трансформера. Она имеет два блока attention с диагональной маской, чтобы на  $i$ -ом шаге трансформер мог делать предсказания не видя текущее значение, а видя только все предыдущие и все последующие. Сначала временной ряд с пропущенными значениями проходит через один блок attention, который заполняет пропуски, после этого полученные предсказания проходят через второй блок attention, чтобы найти более глубокие закономерности временного ряда как во временном измерении, так и в измерении признаков, после этого два полученных предсказания складываются с обучаемыми весами. Архитектуру SAITS можно увидеть на Рисунке 2.2

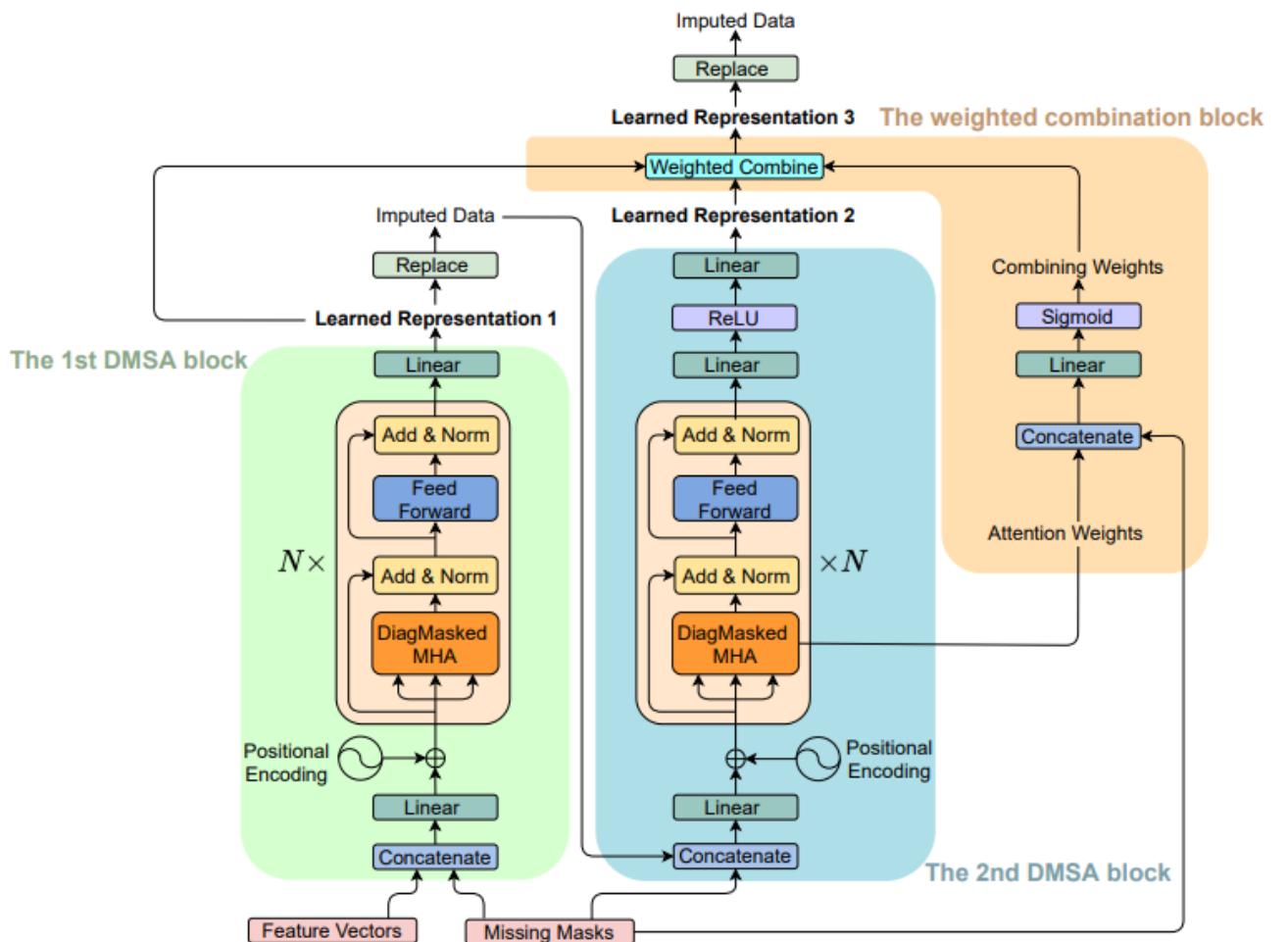


Рис. 2.2: Архитектура SAITS

## 2.5 BRITS

BRITS (Bidirectional Recurrent Imputation for Time Series)[1] в основе своей использует двунаправленную рекуррентную сеть. В отличие от обычной рекуррентной сети, в BRITS

используются маски пропущенных значений  $m_t$ , чтобы вычислять функцию потерь только на неизвестных переменных временного шага, а также временной коэффициент затухания  $\gamma_t = \exp\{-\max(0, \mathbf{W}_\gamma \delta_t + \mathbf{b}_\gamma)\}$ , где  $\delta_t$  — вектор, показывающий для каждой переменной сколько шагов назад было последнее известное её значение, это помогает понимать недостающие закономерности во временном ряду, что имеет решающее значение для расчета. Итоговая функция пересчета скрытых состояний и предсказания значений выглядит так.

$$\begin{aligned}\hat{\mathbf{x}}_t &= \mathbf{W}_x \mathbf{h}_{t-1} + \mathbf{b}_x, \\ \mathbf{x}_t^c &= \mathbf{m}_t \odot \mathbf{x}_t + (1 - \mathbf{m}_t) \odot \hat{\mathbf{x}}_t, \\ \gamma_t &= \exp\{-\max(0, \mathbf{W}_\gamma \delta_t + \mathbf{b}_\gamma)\}, \\ \mathbf{h}_t &= \sigma(\mathbf{W}_h [\mathbf{h}_{t-1} \odot \gamma_t] + \mathbf{U}_h [\mathbf{x}_t^c \circ \mathbf{m}_t] + \mathbf{b}_h), \\ \ell_t &= \langle \mathbf{m}_t, \mathcal{L}_e(\mathbf{x}_t, \hat{\mathbf{x}}_t) \rangle.\end{aligned}\tag{2}$$

## 2.6 Другие сравнительные исследования

В работе [11] приведено подробное описание имеющихся методов прогнозирования временных рядов и заполнения пропусков в них с использованием диффузионных моделей. В ней мы можем ознакомиться с CSDI и другими диффузионными моделями, которые не будут рассматриваться в данной работе. Это обзорное исследование приводит качественный анализ и сравнение методов и принципов работы диффузионных моделей в работе с временными рядами. Основным недостатком данного исследования является отсутствие непосредственных численных экспериментов с реальными данными. В свете отсутствия полноценного обзора и анализа результатов моделей для заполнения пропусков на реальных данных, и предлагается данная работа.

## 3 Постановка задачи заполнения временных рядов

В реальной жизни мы постоянно сталкиваемся с проблемой отсутствующих данных. При сборе информации о временных рядах условия сбора могут меняться, что создает препятствия для получения полной картины. Кроме того, неисправности датчиков и человеческие ошибки могут привести к потере исторических данных. Недостающие значения в данных временных рядов часто негативно влияют на точность анализа и прогнозирования, поскольку отсутствие части информации усложняет обобщение данных.

Формально задача состоит в следующем. Пусть у нас есть временной ряд  $X = \{x_1, \dots, x_T\}$ ,  $x_i \in \mathbb{R}^D$ , состоящий из  $T$  записей, каждая запись это снятие показаний  $D$  вещественных пе-

ременных во времени. Такой временной ряд можно представить матрицей  $X \in \mathbb{R}^{D \times T}$ . В отличие от задачи прогнозирования, где мы считаем матрицу  $X$  известной, в задаче заполнения пропусков часть матрицы содержит пропущенные значения, которые необходимо восстановить.

## 4 Условия экспериментов

Для сравнения разных способов заполнения пропусков первоочередно необходимо усвоиться о наборах данных и условиях, в которых модели будут тестироваться. Для того, чтобы результаты были более общими и точными, необходимо, чтобы данные, на которых будет проводиться тестирование моделей, были разнородными, то есть разными по природе. Для этого были взяты наборы данных из библиотеки PyPOTS[3].

**PhysioNet Mortality Prediction Challenge 2012 (PhysioNet-2012)**[12]. Из набора данных PhysioNet-2012 по прогнозированию смертности взяты 5000 многомерных клинических временных рядов, собранных у пациентов в отделении реанимации и интенсивной терапии. Каждый параметр регистрируется в течение первых 48 часов после поступления в отделение интенсивной терапии. В зависимости от состояния пациентов измеряется до 37 переменных, например, температура, частота сердечных сокращений, артериальное давление. Данные измерений могут собираться с регулярными интервалами (ежечасно или ежедневно), а также могут записываться с нерегулярными интервалами (собираются только по мере необходимости). Не все переменные доступны во всех выборках. Важно обратить внимание, что этот набор данных очень разрежен и в общей сложности содержит 80% пропущенных значений.

**Beijing Multi-Site Air-Quality(Air-Quality)** [17]. Этот набор данных о качестве воздуха включает ежечасные данные о загрязнителях воздуха с 12 метеорологических станций в Пекине. Данные собраны с 2013/03/01 по 2017/02/28 (в общей сложности за 48 месяцев). Для каждого участка мониторинга измеряются 11 переменных непрерывного временного ряда (например, PM2.5, PM10, SO2), но для данного эксперимента, согласно предыдущим работам[1, 4] были взяты только данные PM2.5. В общей сложности в этом наборе данных отсутствует 1.6% значений. Чтобы сгенерировать выборки данных временных рядов, мы выбираем данные за каждые 24 часа, т.е. за каждые 24 последовательных шага, в качестве одной выборки.

**Electricity Load Diagrams (ELD)**[16]. Это еще один широко используемый общедоступный набор данных, он содержит данные о потреблении электроэнергии (в кВт/ч),

собираемые от 370 клиентов каждые 15 минут, и в нем нет пропущенных данных. Период действия этого набора данных - с 2011/01/01 по 2014/12/31 (в общей сложности 48 месяцев). Каждые 96 последовательных шагов(одни полные сутки) выбираются в качестве выборка для генерации данных временных рядов для обучения модели. В данных отсутствуют пропуски.

**Electricity Transformer Temperature (ETT)**[18]. Этот набор данных собран по электрическим трансформаторам за два года с 2016/07/01 по 2018/06/26. Здесь мы выбираем набор данных для почасовой версии(собранной каждый час) выборки без пропусков значений. Каждая выборка содержит семь характеристик, включая температуру масла и шесть различных типов внешних силовых нагрузок.

Для экспериментов в каждом датасете первые 10% записей использовались, как тестовые записи. Из них случайно удалялись 10/50/90% имеющихся значений, которые использовались, как тестовые данные для моделирования случайных пропусков во временных рядах. Модели обучались на оставшихся данных в течение 200 эпох с помощью оптимизатора Adam, параметры всех моделей и расписание learning rate для Adam были взяты из оригинальных статей. Расчеты производились на суперкомпьютере НИУ ВШЭ "сHARISMa" на видеокарте NVIDIA V100 в течение суток

В качестве моделей были взяты BRITS[1], SAITS[4] и GP-VAE[5], описанные выше, и их реализация из библиотеки PyPOTS.

Для сравнения использовались метрики CRPS[7], NLPD[6], PICP[13], а также

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_{true} - x_{pred}|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{true} - x_{pred})^2}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|x_{true} - x_{pred}|}{\max(\epsilon, |x_{true}|)}$$

где  $n$  — количество наблюдений в выборке,  $x_{true}$  — истинное значение временного ряда,  $x_{pred}$  — предсказанное моделью значение временного ряда.

## 5 Результаты

В таблицах ниже можно увидеть метрики, полученные в результате экспериментов и графики заполнений пропусков. Линией отмечены предсказания, красными точками отмечены данные из обучающей выборки, синими крестами отмечены данные тестовой выборки. Для CSDI предсказания для каждого пропущенного значения генерировалось по 30 раз, линией отмечены медиана предсказаний, а закрашена зона между 5% и 95% квантилем распределения предсказаний.

	10% пропусков				50% пропусков			
	MAE	RMSE	MAPE	CRPS	MAE	RMSE	MAPE	CRPS
CSDI	<b>0.204</b>	<b>0.480</b>	<b>0.292</b>	0.507	0.302	0.660	0.435	0.750
SAITS	0.214	0.481	0.307	<b>0.307</b>	<b>0.290</b>	<b>0.628</b>	<b>0.418</b>	<b>0.418</b>
BRITS	0.231	0.546	0.332	0.332	0.316	0.663	0.455	0.455
GP VAE	0.601	0.863	0.863	0.863				

	90% пропусков			
	MAE	RMSE	MAPE	CRPS
CSDI	<b>0.204</b>	<b>0.480</b>	<b>0.292</b>	0.507
SAITS	0.214	0.481	0.307	<b>0.307</b>
BRITS	0.231	0.546	0.332	0.332
GP VAE	0.601	0.863	0.863	0.863

Таблица 5.1: Результаты метрик для датасета PhysioNet (все метрики чем меньше, тем лучше)

	10% пропусков				50% пропусков			
	MAE	RMSE	MAPE	CRPS	MAE	RMSE	MAPE	CRPS
CSDI	17.210	32.907	0.202	0.370	25.733	38.090	0.301	0.446
SAITS	19.263	31.315	0.227	0.227	29.228	42.117	0.468	0.468
BRITS	<b>12.690</b>	<b>22.256</b>	<b>0.149</b>	<b>0.149</b>	<b>16.251</b>	30.443	<b>0.190</b>	<b>0.190</b>
GP VAE	17.205	27.829	0.202	0.202	17.925	<b>28.625</b>	0.218	0.218

	90% пропусков			
	MAE	RMSE	MAPE	CRPS
CSDI	45.541	57.598	0.532	0.819
SAITS	33.398	51.543	0.486	0.486
BRITS	27.091	49.854	0.317	0.317
GP VAE	<b>23.119</b>	<b>40.110</b>	<b>0.270</b>	<b>0.270</b>

Таблица 5.2: Результаты метрик для датасета Air-Quality (все метрики чем меньше, тем лучше)

В таблице 5.5 приведены метрики NLPD и PICP только для CSDI, поскольку для других методов заполнения они невычислимые, так как нет распределения предсказаний модели.

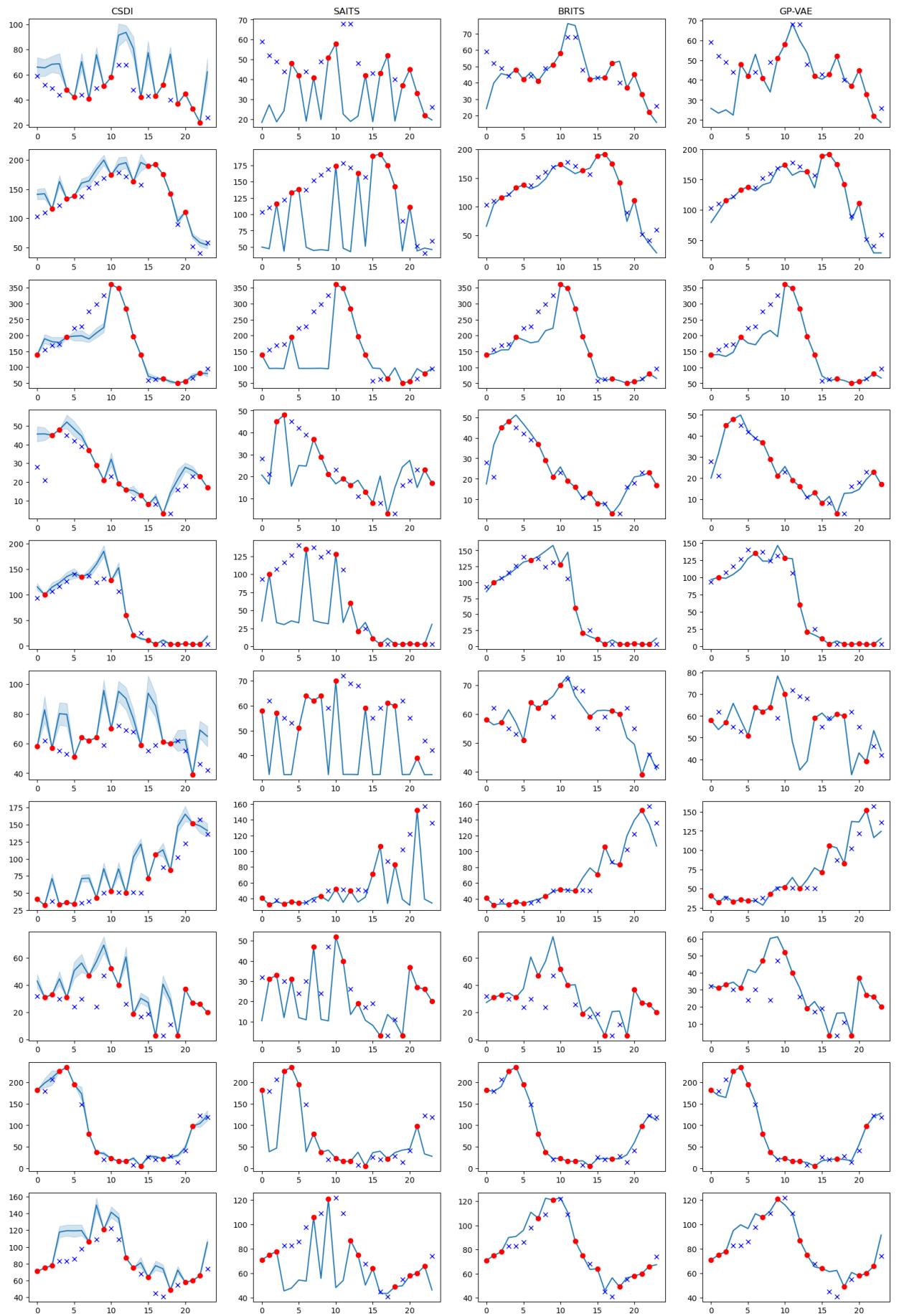


Рис. 5.1: Предсказания для датасета Air-Quality, 50% значений пропущено

	10% пропусков				50% пропусков			
	MAE	RMSE	MAPE	CRPS	MAE	RMSE	MAPE	CRPS
CSDI	<b>0.302</b>	0.601	<b>0.055</b>	0.106	0.800	2.516	0.146	0.234
SAITS	0.306	<b>0.562</b>	0.056	<b>0.056</b>	<b>0.502</b>	<b>0.923</b>	<b>0.092</b>	<b>0.092</b>
BRITS	0.402	1.002	0.073	0.073	0.671	1.410	0.123	0.123
GP VAE	1.039	1.516	0.189	0.189	1.892	2.649	0.344	0.344
90% пропусков								
	MAE	RMSE	MAPE	CRPS				
CSDI	2.510	4.599	0.460	0.665				
SAITS	<b>1.452</b>	<b>2.622</b>	<b>0.266</b>	<b>0.266</b>				
BRITS	2.012	3.587	0.368	0.368				
GP VAE	2.916	4.894	0.534	0.534				

Таблица 5.3: Результаты метрик для датасета ETT (все метрики чем меньше, тем лучше)

	10% пропусков				50% пропусков			
	MAE	RMSE	MAPE	CRPS	MAE	RMSE	MAPE	CRPS
CSDI	10.343	20.034	0.137	0.234	14.889	<b>22.030</b>	0.198	0.276
SAITS	<b>8.900</b>	<b>15.097</b>	<b>0.118</b>	<b>0.118</b>	15.421	26.967	0.205	0.205
BRITS	10.713	18.180	0.141	0.141	<b>13.631</b>	23.790	<b>0.181</b>	<b>0.181</b>
GP VAE	11.092	17.991	0.146	0.146	15.870	28.148	0.211	0.211
90% пропусков								
	MAE	RMSE	MAPE	CRPS				
CSDI	16.327	<b>23.308</b>	0.216	0.328				
SAITS	17.145	30.750	0.228	0.228				
BRITS	17.514	31.282	0.232	0.232				
GP VAE	<b>14.863</b>	25.497	<b>0.197</b>	<b>0.197</b>				

Таблица 5.4: Результаты метрик для датасета ELD (все метрики чем меньше, тем лучше)

## 6 Выводы

На большинстве наборов SAITS показал себя лучше всех, при малом проценте пропусков CSDI в некоторых случаях выигрывает, но при увеличении доли пропусков не может догнать SAITS. Помимо этого SAITS обучался и генерировал предсказания в 3 раза быстрее CSDI. BRITS и GP-VAE показывают худшие результаты по сравнению с SAITS и CSDI, выигрывают они только на датасете Air-Quality. Также нужно отметить, что CRPS у CSDI больше, чем MAPE в отличие от остальных моделей, для которых это одна и та же метрика, мы можем сделать вывод, что распределение предсказаний CSDI не так хорошо

	10% пропусков		50% пропусков		90% пропусков	
	NLPD ↓	PICP↑	NLPD ↓	PICP↑	NLPD ↓	PICP↑
PhysioNet	54.696	0.883	92.944	0.875	133.281	0.691
Air-Quality	12.830	0.914	6.009	0.720	6.703	0.701
ETT	8.499	0.883	1.796	0.857	2.725	0.683
ELD	13.282	0.876	15.274	0.506	46.570	0.650

Таблица 5.5: Результаты метрик для CSDI на различных датасетах

описывает распределение данных, как хотелось бы и оказывается хуже точечного предсказания. Диффузионные модели, и CSDI в их числе — передовые методы генерации, которые продемонстрировали свою исключительную эффективность в различных реальных приложениях, но они сопряжены с большими вычислительными затратами при сопоставимых с трансформерами результатах. Также важно отметить, что для качественного применения CSDI необходимо нормировать данные и сгенерировать не менее нескольких десятков вариантов для заполнения каждого пропуска, поскольку иначе качество предсказаний оказывается хуже SAITS. Учитывая все сказанное, SAITS оказывается самым универсальным и работоспособным методом заполнения пропусков, поскольку работает быстрее и качественнее остальных, диффузионные методы заполнения пропусков во временных рядах требуют дальнейшего изучения и улучшения.

## Список литературы

- [1] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li и Yitan Li. “BRITS: Bidirectional Recurrent Imputation for Time Series”. B: *Neural Information Processing Systems*. 2018.
- [2] Yankun Chen, Jingxuan Liu, Lingyun Peng, Yiqi Wu, Yige Xu и Zhanhao Zhang. “Auto-Encoding Variational Bayes”. B: *Cambridge Explorations in Arts and Sciences* 2 (февр. 2024). DOI: [10.61603/ceas.v2i1.33](https://doi.org/10.61603/ceas.v2i1.33).
- [3] Wenjie Du. “PyPOTS: a Python toolbox for data mining on Partially-Observed Time Series”. B: (2023). DOI: [10.48550/arXiv.2305.18811](https://doi.org/10.48550/arXiv.2305.18811). arXiv: [2305.18811 \[cs.LG\]](https://arxiv.org/abs/2305.18811). URL: <https://arxiv.org/abs/2305.18811>.
- [4] Wenjie Du, David Côté и Yan Liu. “SAITS: Self-attention-based imputation for time series”. B: *Expert Systems with Applications* 219 (июнь 2023), с. 119619. ISSN: 0957-4174.
- [5] Vincent Fortuin, Dmitry Baranchuk, Gunnar Ratsch и Stephan Mandt. “GP-VAE: Deep probabilistic time series imputation”. B: *International conference on artificial intelligence and statistics*. PMLR. 2020, с. 1651–1661.
- [6] Biraja Ghoshal, Allan Tucker, Bal Sanghera и Wai Lup Wong. “Estimating uncertainty in deep learning for reporting confidence to clinicians in medical image segmentation and diseases detection”. B: *Computational Intelligence* 37 (2020), с. 701–734.
- [7] Tilmann Gneiting и Adrian Raftery. “Strictly Proper Scoring Rules, Prediction, and Estimation”. B: *Journal of the American Statistical Association* 102 (март 2007), с. 359—378.
- [8] Jonathan Ho, Ajay Jain и Pieter Abbeel. “Denoising Diffusion Probabilistic Models”. B: *arXiv preprint arxiv:2006.11239* (2020).
- [9] Igor Ilic, Berk Gorgulu, Mucahit Cevik и Mustafa Gokce Baydogan. *Explainable boosted linear regression for time series forecasting*. 2020. arXiv: [2009.09110 \[cs.LG\]](https://arxiv.org/abs/2009.09110).
- [10] Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao и Bryan Catanzaro. “DiffWave: A Versatile Diffusion Model for Audio Synthesis”. B: *International Conference on Learning Representations*. 2021.
- [11] Lequan Lin, Zhengkun Li, Ruikun Li, Xuliang Li и Junbin Gao. “Diffusion Models for Time Series Applications: A Survey”. B: *Frontiers of Information Technology and Electronic Engineering* (2023).

- [12] Ikaro Silva, George Moody, Daniel J Scott, Leo A Celi и Roger G Mark. “Predicting in-hospital mortality of icu patients: The physionet/computing in cardiology challenge 2012”. B: *Computing in cardiology* 39 (2012), c. 245.
- [13] Laurens Sluijterman, Eric Cator и Tom Heskes. *How to Evaluate Uncertainty Estimates in Machine Learning for Regression?* 2023. arXiv: [2106.03395 \[stat.ML\]](https://arxiv.org/abs/2106.03395).
- [14] Samya Tajmouati, Bouazza El Wahbi, Adel Bedoui, Abdallah Abarda и Mohamed Dakkoun. *Applying k-nearest neighbors to time series forecasting : two new approaches.* 2021. arXiv: [2103.14200 \[stat.ME\]](https://arxiv.org/abs/2103.14200).
- [15] Yusuke Tashiro, Jiaming Song, Yang Song и Stefano Ermon. “CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation”. B: *Advances in Neural Information Processing Systems*. Под ред. M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang и J. Wortman Vaughan. Т. 34. Curran Associates, Inc., 2021, c. 24804—24816.
- [16] Artur Trindade. *ElectricityLoadDiagrams20112014*. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C58C86>. 2015.
- [17] Shuyi Zhang, Bin Guo, Anlan Dong, Jing He, Ziping Xu и Song Xi Chen. “Cautionary tales on air-quality improvement in Beijing”. B: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 473 (2017). URL: <https://api.semanticscholar.org/CorpusID:37683936>.
- [18] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong и Wancai Zhang. “Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting”. B: *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*. Т. 35. 12. AAAI Press, 2021, c. 11106—11115.

## А Примеры графиков предсказаний

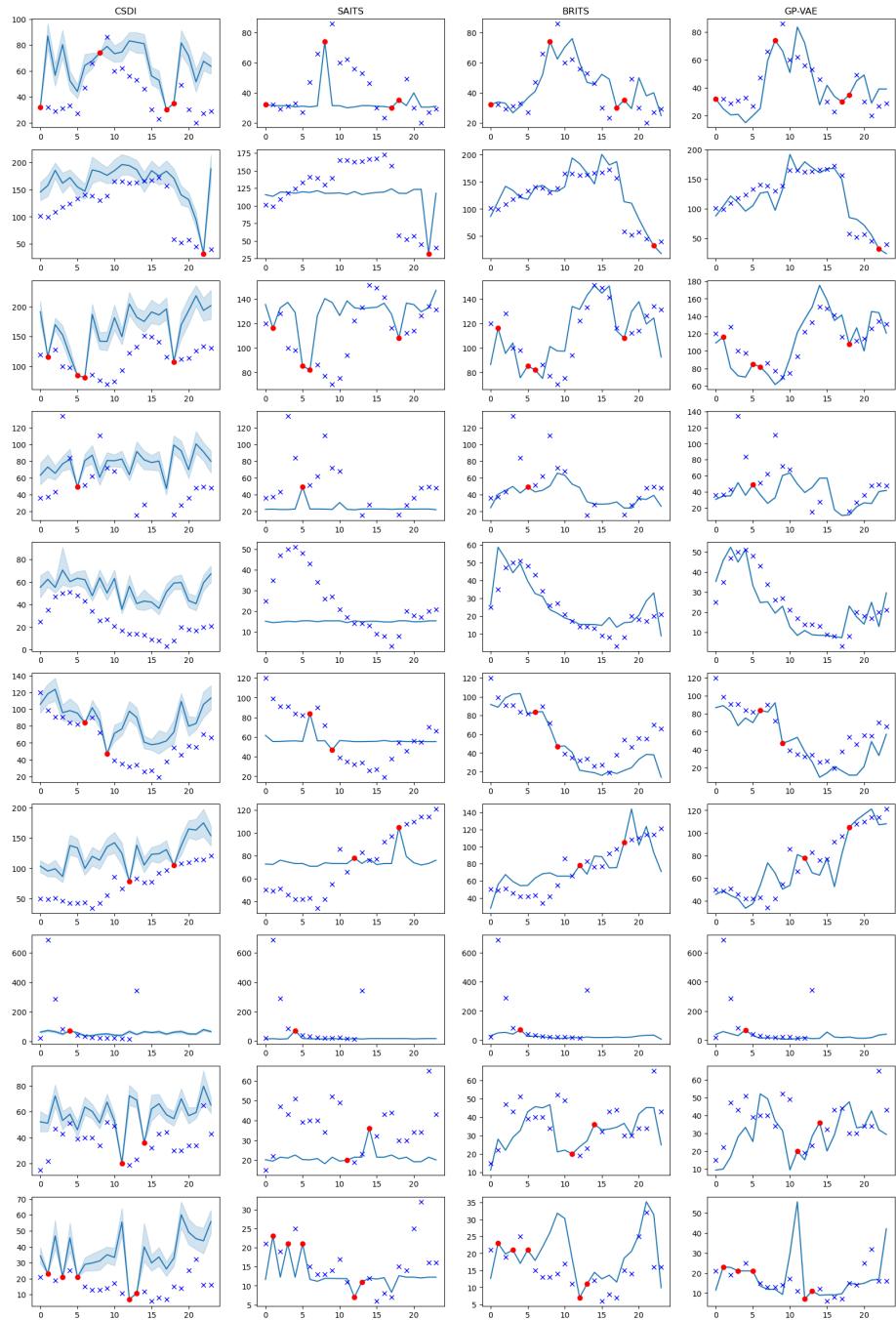


Рис. А.1: Предсказания для датасета Air-Quality, 90% значений пропущено



Рис. А.2: Предсказания для датасета PhysioNet, 50% значений пропущено

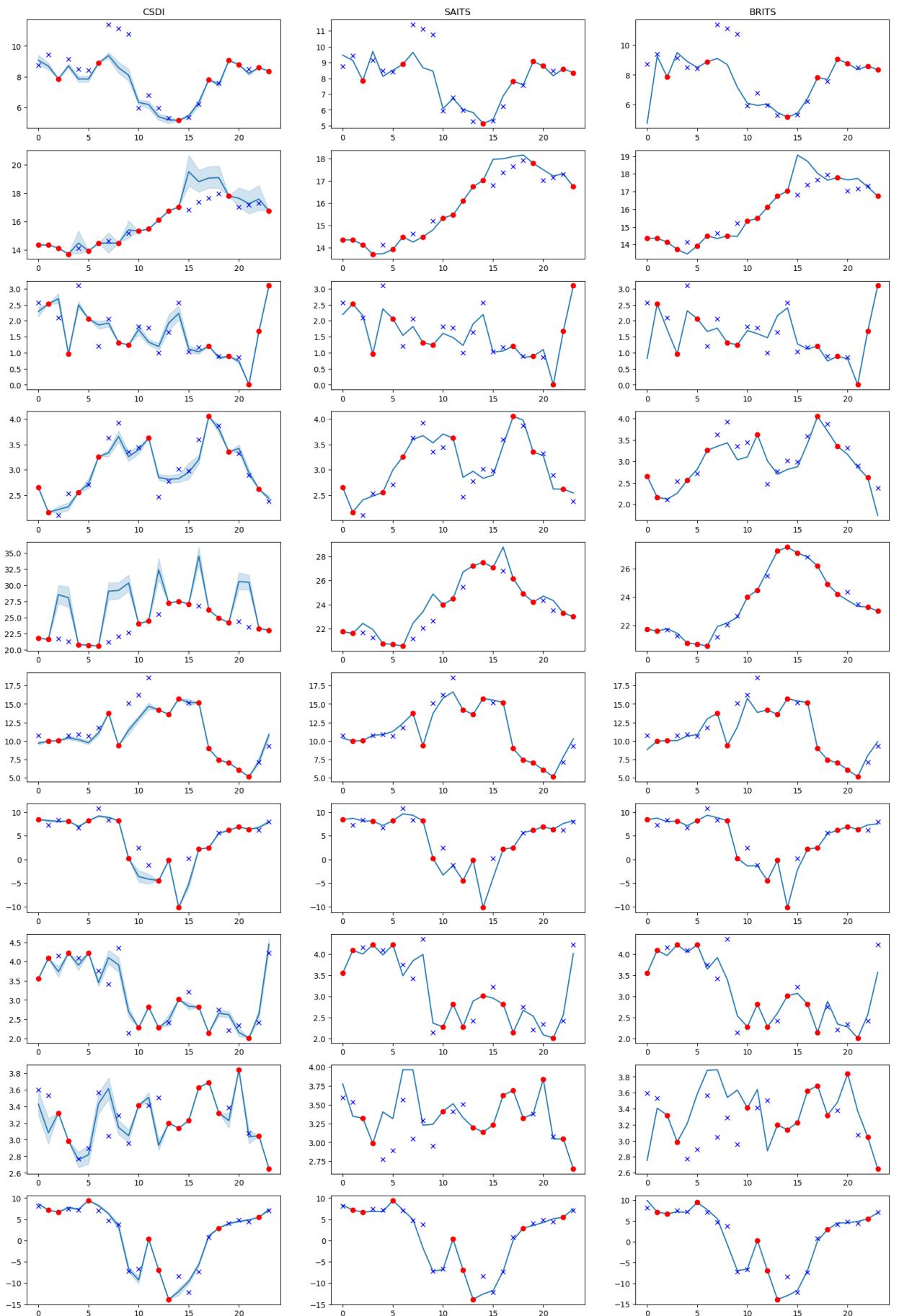


Рис. А.3: Предсказания для датасета ETT, 50% значений пропущено

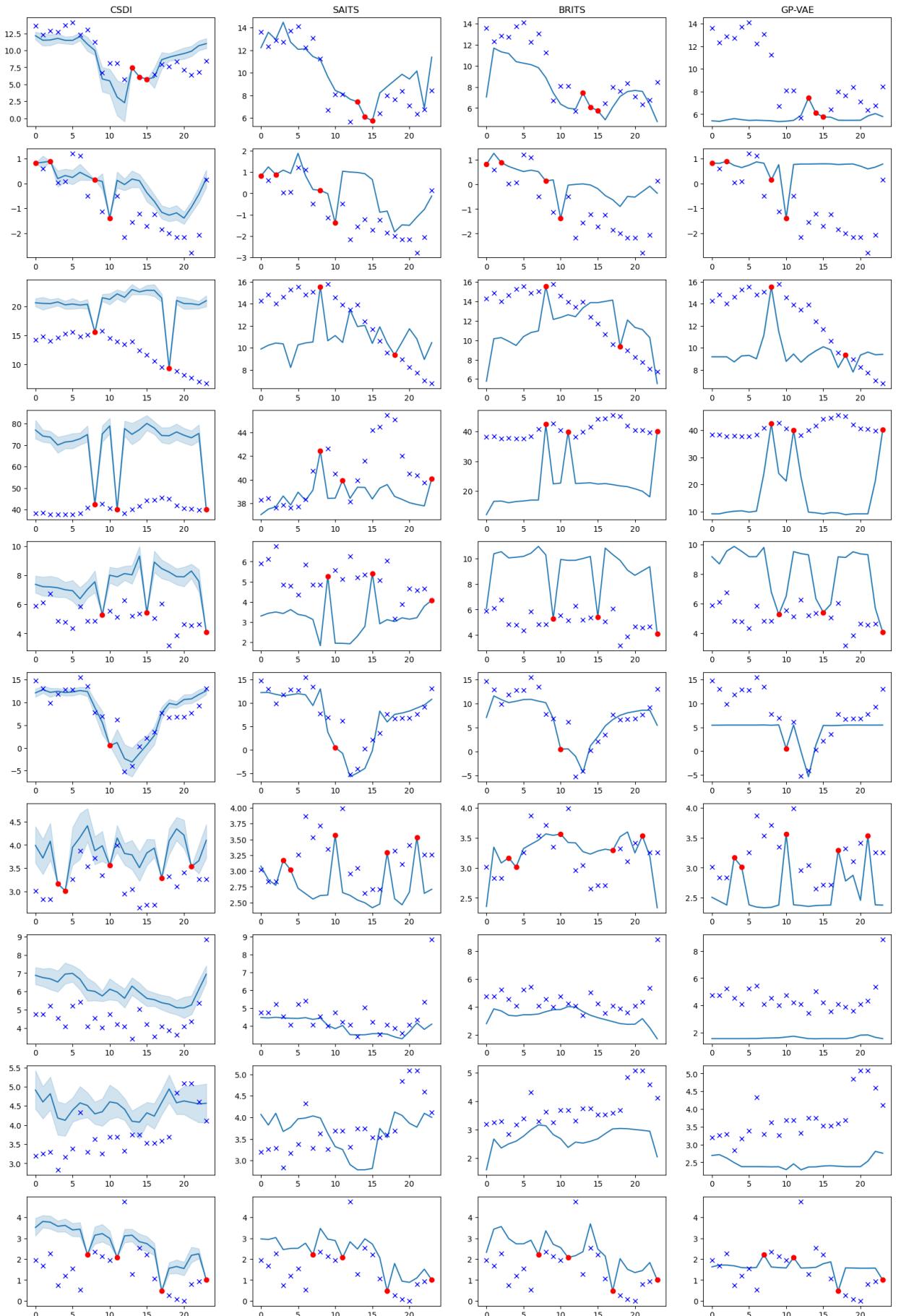


Рис. А.4: Предсказания для датасета ETT, 90% значений пропущено

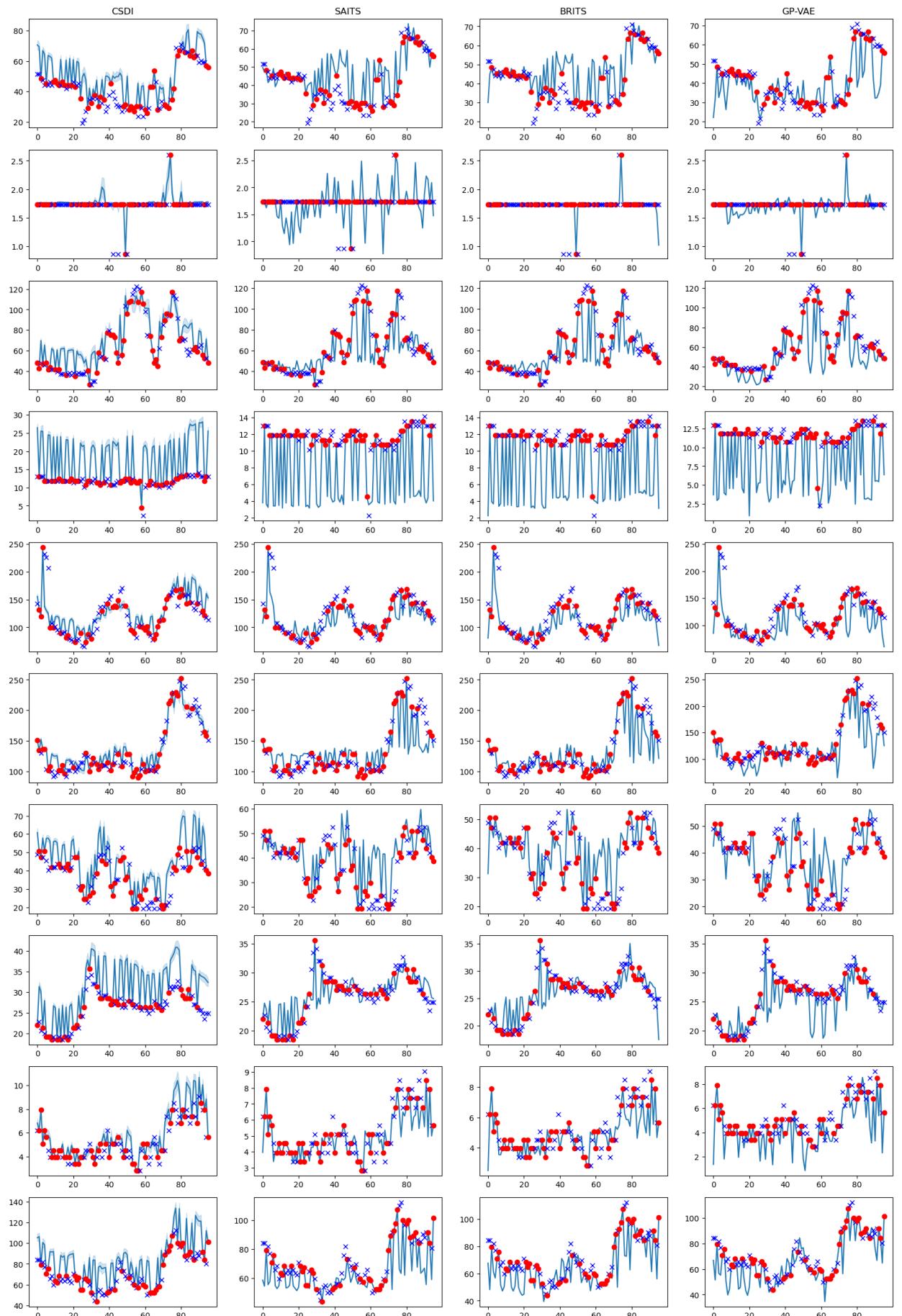


Рис. А.5: Предсказания для датасета ELD, 50% значений пропущено

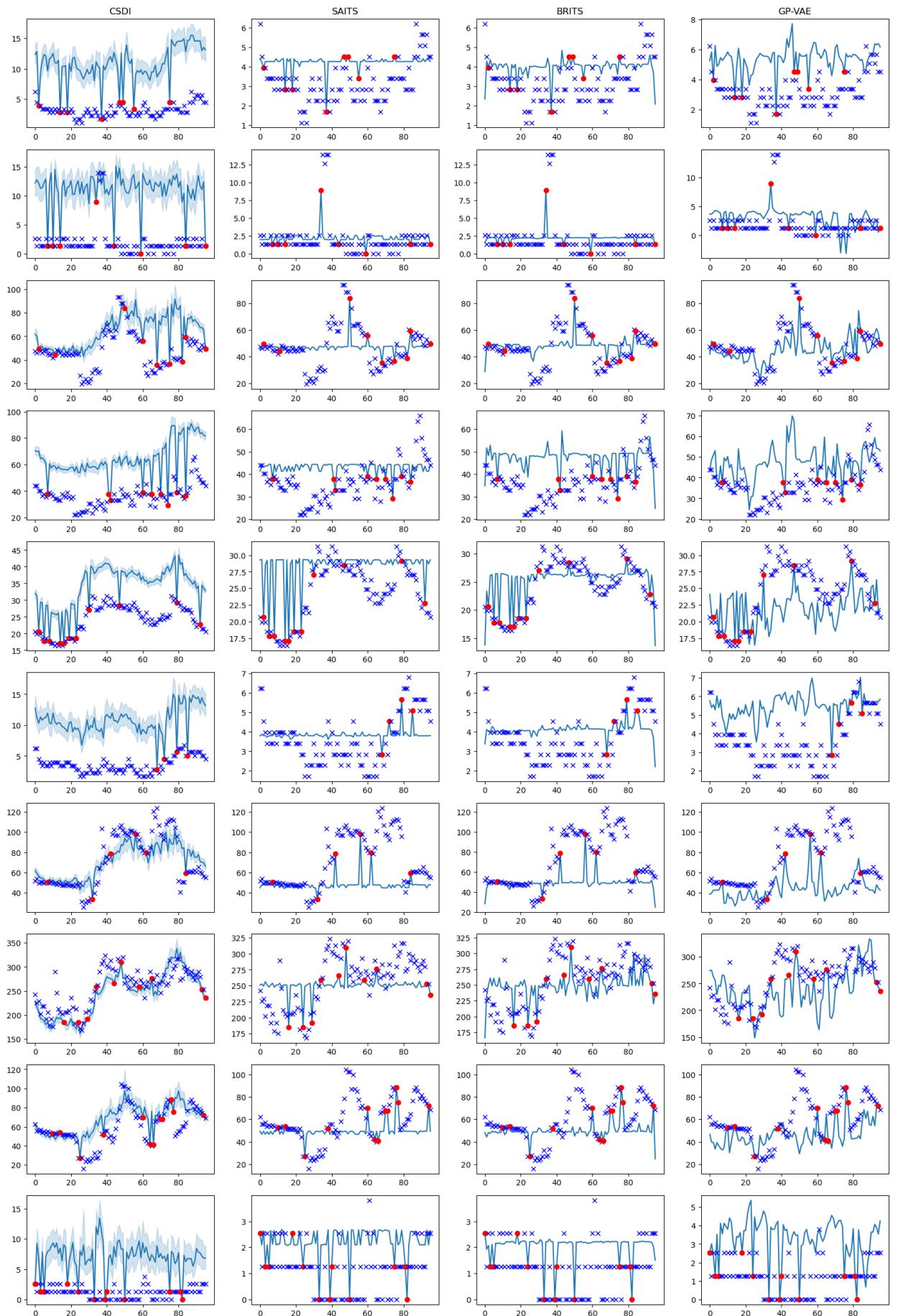


Рис. А.6: Предсказания для датасета ELD, 90% значений пропущено