УДК 004.852:519.257:338.43(575.2)
DOI: 10.36979/1694-500X-2026-26-8-65-73

ВАРИАЦИОННЫЙ АВТОЭНКОДЕР ДЛЯ ИМПУТАЦИИ ПАНЕЛИ АГРОЭКОНОМИЧЕСКИХ ДАННЫХ КЫРГЫЗСКОЙ РЕСПУБЛИКИ: РЕЗУЛЬТАТЫ ОБУЧЕНИЯ, SELF-MASKING-ВАЛИДАЦИЯ И ДИАГНОСТИКА ПО ФАКТИЧЕСКОМУ ПРОГОНУ ПАЙПЛАЙНА

Ж.Б. Мамадалиева, А.Д. Картанова, Ж.Р. Сарыпбекова

Аннотация. Показаны результаты фактического запуска пайплайна VAE-импутации на панели агроэкономических данных Кыргызской Республики в двух вариантах – без аугментации, обучающей выборки, и с полной аугментацией, – продолжая методологическое описание, данное в статье. Приводятся точные показатели пайплайна: объём и структура признакового пространства, доля и структура пропусков, кривые обучения ELBO, влияние аугментации на корреляционную структуру данных и на итоговое качество импутации, оценка неопределённости методом Monte Carlo Dropout на реально пропущенном наблюдении и self-masking-валидация на трёх контрольных культурах (хлопок, сахарная свёкла, пшеница) в сравнении с методом среднего, KNN и MICE. Ключевой результат: аугментация обучающей выборки в этом прогоне не улучшила, а ухудшила и валидационный loss VAE (0,0616 → 0,1449), и точность self-masking восстановления VAE на всех трёх контрольных культурах; MICE при этом устойчиво превосходит VAE по точности восстановления на всех трёх культурах в обоих вариантах пайплайна. Результат сопровождается диаграммами, полученными непосредственно из данного прогона: коррелограммы до/после аугментации, кривые обучения, гистограмма MC Dropout, box-plot абсолютных ошибок self-masking, а также сравнение истинных и VAE-импутированных значений урожайности пшеницы на реально пропущенных наблюдениях.

Ключевые слова: вариационный автоэнкодер; импутация пропущенных данных; self-masking; Monte Carlo Dropout; аугментация данных; MICE; агроэкономическая панель; Кыргызская Республика.

КЫРГЫЗ РЕСПУБЛИКАСЫНЫН АГРОЭКОНОМИКАЛЫК ПАНЕЛИН ТОЛУКТОО ҮЧҮН ВАРИАЦИЯЛЫК АВТОЭНКОДЕР: ОКУТУУНУН ЖЫЙЫНТЫКТАРЫ, SELF-MASKING ВАЛИДАЦИЯСЫ ЖАНА ПАЙПЛАЙНДЫ РЕАЛДУУ ИШТЕТҮҮДӨН АЛЫНГАН ДИАГНОСТИКА

Ж.Б. Мамадалиева, А.Д. Картанова, Ж.Р. Сарыпбекова

Аннотация. Бул макалада Кыргыз Республикасынын агроэкономикалык берилиштер панелине VAE ыкмасы менен берилиштерди толуктоо (импутация) пайплайнынын реалдуу иштешинин жыйынтыктары берилет. Пайплайн эки вариантта каралат: окутуу үлгүсүн аугментациясыз жана толук аугментация менен, бул макалада берилген методологиялык сүрөттөөнү толуктап турат. Пайплайндын так мүнөздөмөлөрү келтирилет: белгилер мейкиндигинин көлөмү жана түзүмү, бош маанилердин үлүшү жана түзүмү, ELBO окутуу ийри сызыктары, аугментациянын берилиштердин корреляциялык структурасына жана толуктоонун акыркы сапатына тийгизген таасири; чыныгы жетишпеген байкоолордо Monte Carlo Dropout жолу менен белгисиздикти баалоо жана үч контролдук өсүмдүккө (пахта, кант кызылы, буудай) жасалган self-masking валидациясы орточо мааниге, KNN жана MICE ыкмаларына салыштырмалуу берилет. Негизги жыйынтык: бул экспериментте окутуу үлгүсүн аугментациялоо VAE үчүн валидациялык loss маанисин (0,0616 → 0,1449) да, үч контролдук өсүмдүктүн self-masking негизиндеги калыбына келтирүү тактыгын да жакшыртпастан, тескерисинче начарлатты; ошол эле учурда MICE эки вариантта тең үч өсүмдүктү толуктоодо VAE ыкмасынан тактыгы боюнча туруктуу түрдө үстөмдүк кылат. Натыйжалар түз ушул иштетүүдөн алынган диаграммалар менен коштолгон: аугментацияга чейинки жана кийинки коррелограмма, ELло окутуу ийри сызыктары, MC Dropout гистограммасы, self-masking абсолюттук каталардын box-plot диаграммалары, ошондой эле буудай түшүмүнүн чыныгы жана VAE менен толукталып алынган маанилеринин салыштыруу графиги көрсөтүлгөн.

Түйүндү сөздөр: вариациялык автоэнкодер; жетишпеген маалыматтарды толуктоо; self-masking; Monte Carlo Dropout; маалыматтарды аугментациялоо; MICE; агроэкономикалык панель; Кыргыз Республикасы.

VARIATIONAL AUTOENCODER FOR IMPUTING THE PANEL OF AGRO-ECONOMIC DATA OF THE KYRGYZ REPUBLIC: TRAINING RESULTS, SELF-MASKING VALIDATION, AND DIAGNOSTICS FROM THE ACTUAL PIPELINE RUN

Zh.B. Mamadalieva, A.D. Kartanova, Zh.R. Sarypbekova

Abstract. This article presents the results of the actual execution of the VAE-based imputation pipeline on the agro-economic data panel of the Kyrgyz Republic under two configurations: without training-set augmentation and with full augmentation, thereby continuing the methodological description provided. The paper reports detailed pipeline characteristics, including the size and structure of the feature space, the proportion and pattern of missing values, ELBO learning curves, the impact of augmentation on the correlation structure and the resulting imputation quality, uncertainty estimation using the Monte Carlo Dropout method for truly missing observations, and self-masking validation for three control crops: cotton, sugar beet, and wheat, in comparison with mean imputation, KNN, and MICE. The key result is that, in this pipeline run, training-set augmentation did not improve but instead worsened both the VAE validation loss (0,0616 → 0,1449) and the self-masking reconstruction accuracy of the VAE for all three control crops. At the same time, MICE consistently outperformed the VAE in reconstruction accuracy for all three crops under both pipeline configurations. These findings are supported by diagrams generated directly from the actual run, including correlation matrices before and after augmentation, ELBO learning curves, a Monte Carlo Dropout histogram, box plots of self-masking absolute errors, and a comparison of true and VAE-imputed wheat yield values for actually missing observations.

Keywords: variational autoencoder; missing-data imputation; self-masking; Monte Carlo Dropout; data augmentation; MICE; agro-economic panel; Kyrgyz Republic.

Введение. Настоящая статья, в отличие от методологического описания в статье [1], отчитывается по фактическому исполнению пайплайна VAE-импутации на актуальной версии панели [2, 3]: приводятся числа, кривые обучения и диагностические диаграммы одного конкретного запуска (315 наблюдений «регион – год»), а не общее описание архитектуры. Архитектура маскированного энкодера и декодера, агрономически взвешенная функция потерь ELBO и пятикомпонентная схема аугментации обучающей выборки подробно описаны в статье [1]; здесь эта методология применяется к конкретному прогону и проверяется количественно. Пайплайн выполнен дважды – как запуск без аугментации и запуск с аугментацией, – что позволяет напрямую сравнить эффект аугментации на loss обучения и на self-masking точность.

Данные и признаковое пространство. Загружено 315 строк и 107 столбцов (девять регионов, 1991–2025 гг.). Признаков до фильтрации – 113; после удаления признаков с долей пропусков выше 80 % (4 признака) остаётся 109 – все они проходят фильтр по константности без потерь. Признаковое пространство группируется на семь блоков (таблица 1).

Таблица 1 – Признаковое пространство

Группа признаков	Число признаков
yield (урожайность)	15
climate (климат)	17
economic (экономика)	56
areas (площади)	15
machinery (техника)	4
prices_log (цены, логарифм)	6
flags (служебные флаги)	3
Итого	109

Совокупный объём пропусков в сырых данных – 8535 ячеек (25,6 % от всех значений); наиболее разреженные признаки – данные о сельскохозяйственной технике (пропуски 82,9 %) и площадях/производстве хлопка и сахарной свёклы (65–71 %), что согласуется с их публикацией лишь за отдельные годы, отмеченным ещё в [2]. После предобработки и до нормализации доля пропусков составляет 7410 ячеек (21,6 % от 109 признаков × 315 наблюдений); вход энкодера при F = 109 имеет размерность 2F = 218 (см. архитектуру в статье [1]).

Обучение без аугментации. Разбиение выполняется на уровне исходных наблюдений (во избежание утечки данных между train и val): из 315 оригиналов 252 отнесены к train и 63 – к val. Обучение идёт 300 эпох; зафиксированы контрольные точки: эпоха 1 – train 0,3855 / val 0,2674; эпоха 100 – train 0,0847 / val 0,0761; эпоха 200 – train 0,0756 / val 0,0700; эпоха 300 – train 0,0710 / val 0,0657. Лучший val loss за прогон – 0,0616, что соответствует высокому качеству реконструкции по ориентирам, принятым для данной задачи.

Аугментация обучающей выборки в данном прогоне. Пять приёмов аугментации, описанные в статье [1] (гауссов шум, целевые климатические сценарии, региональный MixUp, bootstrap, случайное маскирование), дают в этом прогоне следующий прирост наблюдений: гауссов шум + 1575, климатические сценарии + 945, региональный MixUp + 208, bootstrap + 630, случайное маскирование + 315 – итого 3988 кандидатов до обрезки, из которых случайной подвыборкой при обязательном сохранении всех 315 исходных строк формируется финальный пул из 2500 наблюдений (7,9-кратный прирост, что соответствует коэффициенту, приведённому в статье [1]).

Контроль искажения корреляционной структуры (рисунок 1) по восьми отобранным признакам (урожайность пшеницы, хлопка и сахарной свёклы, коэффициент Селянинова, средняя вегетационная температура, сумма осадков вегетационного периода, число морозных дней, ценовой индекс фруктов и ягод) даёт среднее абсолютное отклонение корреляций 0,094 – ниже порога 0,1, то есть аугментация формально не искажает связи между признаками сверх допустимого предела, хотя и близко подходит к нему.

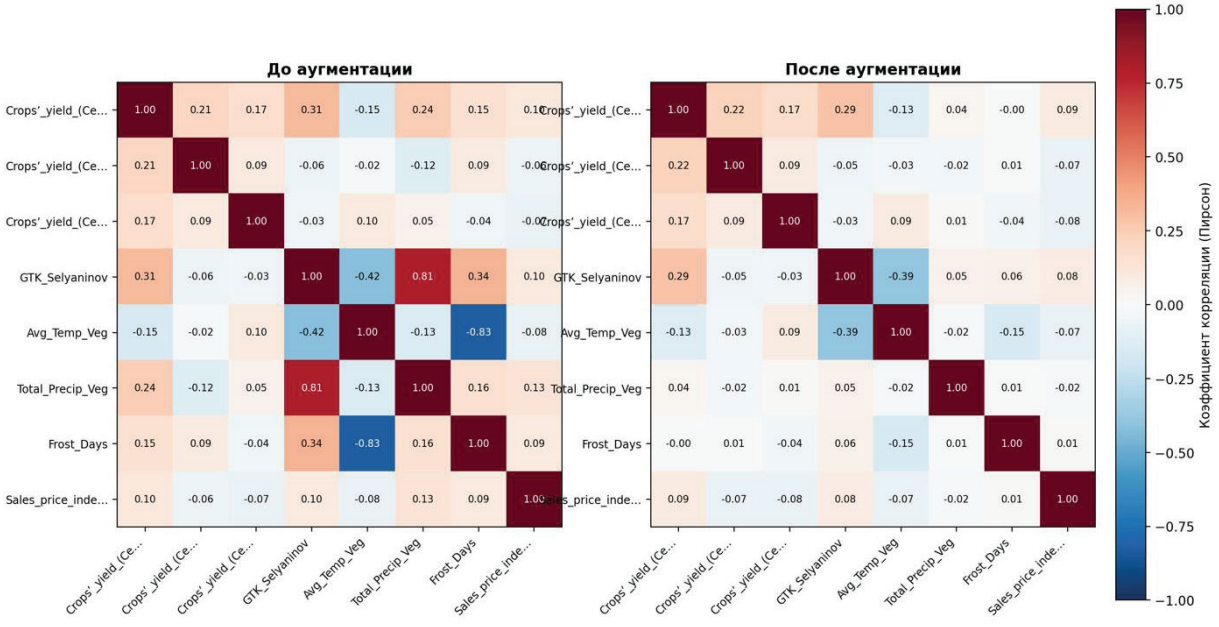


Рисунок 1 – Кореллограммы восьми ключевых признаков до и после аугментации (данный прогон): среднее абсолютное отклонение корреляций – 0,094

Обучение с аугментацией и влияние на loss. Из 2500 наблюдений на уровне оригиналов 2000 отнесены к train и 500 – к val; после фильтрации по признакам, актуальным для self-masking и валидным маскам, остаётся 1655 train и 344 val. Обучение также идёт 300 эпох: эпоха 1 – train 2,1252 / val 1,7897; эпоха 100 – train 0,2356 / val 0,1892; эпоха 200 – train 0,1944 / val 0,1640; эпоха 300 – train 0,1847 / val 0,1472. Лучший val loss – 0,1449 – формально приемлемый показатель, но значительно уступающий по абсолютной величине варианту без аугментации (рисунок 2).

Итоговое сравнение однозначно: минимальный val loss без аугментации составил 0,0616, тогда как при обучении на аугментированном пуле – 0,1449, то есть аугментация не улучшила, а более чем удвоила величину функции потерь на валидации. Это отчасти ожидаемо: без аугментации модель обучается и валидируется на 315 реальных наблюдениях с относительно простой по объёму целью; с аугментацией – на существенно более разнообразном и намеренно зашумлённом пуле из 2500 наблюдений, где сама неопределённость данных выше по построению, а не по качеству модели. Тем не менее для итоговой оценки пригодности аугментации важна не абсолютная величина loss на разных по составу выборках, а её влияние на self-masking точность (раздел 7), где сравнение проводится на одних и тех же контрольных значениях [4].

Оценка неопределённости: Monte Carlo Dropout. MC Dropout выполняется в 100 проходов (рисунок 3). В обоих вариантах пайплайна рассчитана агрегированная диагностика стандартного отклонения по всем импутированным ячейкам (в нормализованном пространстве): без аугментации – минимум 0,0000, среднее 0,0386, медиана 0,0000, максимум 0,8212; с аугментацией – минимум 0,0000, среднее – 0,0207, медиана – 0,0000, максимум – 0,4441, то есть в среднем модель, обученная на аугментированном пуле, даёт более узкие (более уверенные) интервалы неопределённости. Дополнительно для одного реально пропущенного наблюдения – урожайности хлопка в Баткене за 1991 год, выбранного пайплайном автоматически в варианте с аугментацией, – сохранены все 100 индивидуальных прогнозов MC Dropout для детальной иллюстрации формы распределения [5, 6].

Self-masking валидация. Self-masking выполнена на трёх контрольных культурах – хлопок, сахарная свёкла, пшеница – для обоих вариантов пайплайна (таблицы 2, 3) (методология процедуры описана в статье [1]). Приводимые ниже значения – RMSE (ц/га) восстановления искусственно скрытых значений; n – число self-masking наблюдений по культуре.

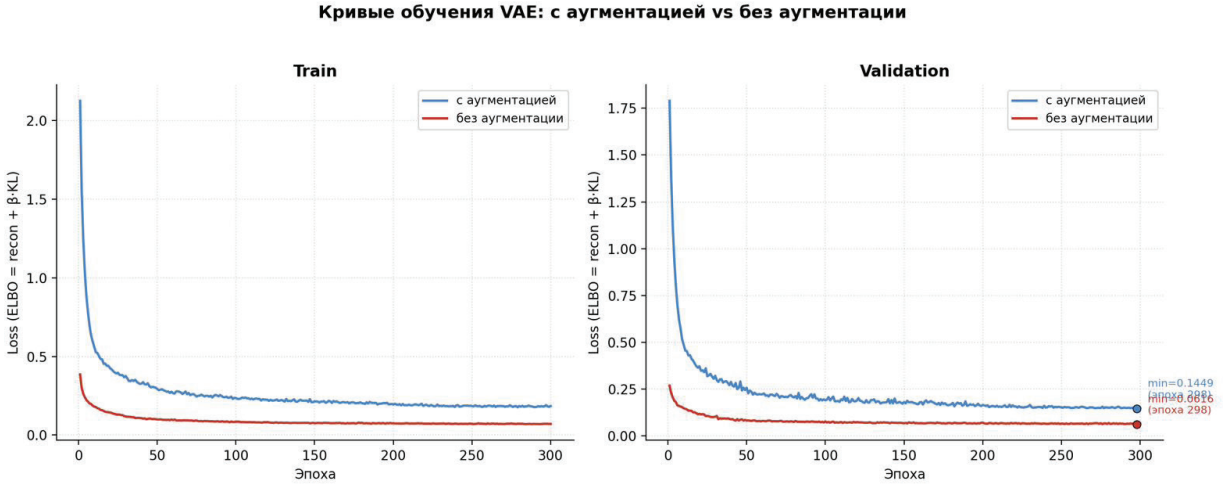


Рисунок 2 – Кривые обучения (ELBO: train/val) без аугментации и с аугментацией. Без аугментации loss выходит на существенно более низкий уровень (val ≈ 0,062) по сравнению с аугментированным вариантом (val ≈ 0,145)

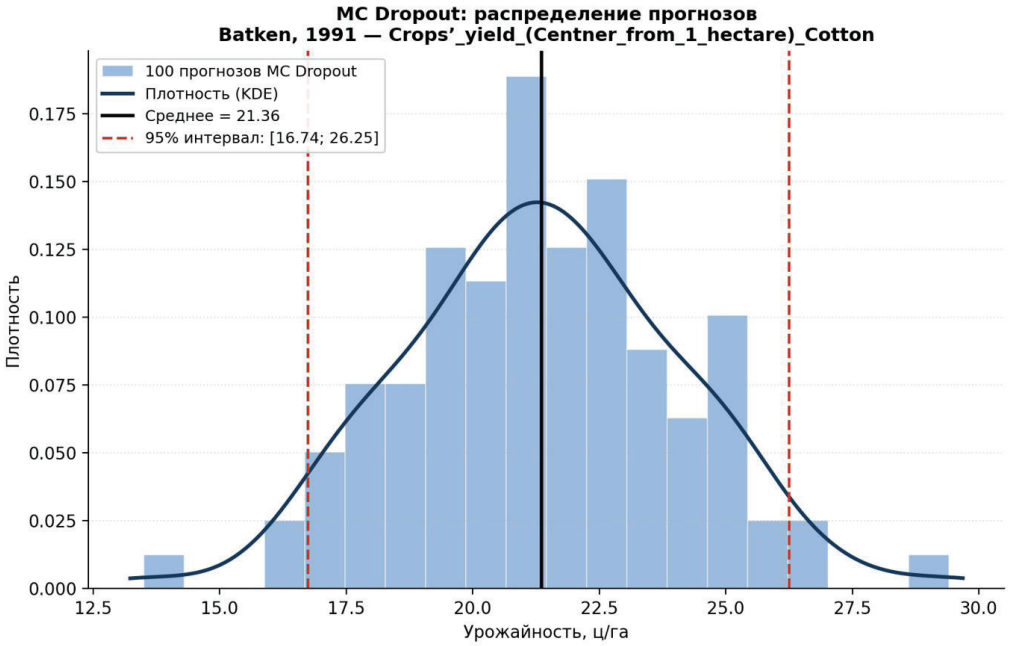


Рисунок 3 – Гистограмма 100 прогнозов MC Dropout для реально пропущенного значения (урожайность хлопка, Баткен, 1991 год; вариант с аугментацией): среднее – 21,36 ц/га, 95%-й интервал – (16,74; 26,25)

Таблица 2 – Результаты обучение без аугментации

Культура (n)	Mean	KNN-5	MICE	VAE
Хлопок (n = 22)	9,26	10,22	3,60	6,34
Сахарная свёкла (n = 18)	133,00	88,79	75,60	94,53
Пшеница (n = 58)	4,70	4,38	3,43	3,64

Таблица 3 – Результаты обучения с аугментацией

Культура (n)	Mean	KNN-5	MICE	VAE
Хлопок (n = 22)	9,26	10,22	3,60	8,49
Сахарная свёкла (n = 18)	133,00	88,79	75,60	122,62
Пшеница (n = 58)	4,70	4,38	3,43	4,03

Результат содержательно однозначен и совпадает по направлению для обеих версий пайплайна (рисунок 4): MICE устойчиво показывает наименьшую (лучшую) RMSE на всех трёх культурах – хлопок (3,60), сахарная свёкла (75,60), пшеница (3,43), – превосходя VAE в обоих вариантах. VAE не является стабильно вторым по качеству методом: на хлопке и пшенице (без аугментации) VAE обгоняет и Mean, и KNN-5, но на сахарной свёкле (без аугментации) VAE (94,53) уступает даже KNN-5 (88,79). Наиболее заметный эффект – аугментация ухудшает точность VAE на всех трёх культурах без исключения: хлопок 6,34 → 8,49 (+34%), сахарная свёкла 94,53 → 122,62 (+30%), пшеница 3,64 → 4,03 (+11%). Показатели Mean, KNN-5 и MICE не меняются между вариантами, поскольку эти методы не используют VAE-модель и не зависят от аугментации [7, 8].

Иллюстрация фактической импутации: пшеница. В отличие от self-masking (искусственного скрытия известных значений для проверки точности), приведённый ниже рисунок показывает результат работы модели на реально пропущенных ячейках панели по показателю урожайности пшеницы в трёх регионах – Чуй, Ош и город Ош. В регионах Чуй и Ош за 1991–2025 гг. пропусков по данному показателю не было, поэтому кривые истинных и VAE-восстановленных значений на всём периоде совпадают. В городе Ош пропущено 5 наблюдений; восстановленные VAE значения (отмечены крестиками) попадают в диапазон соседних по времени точек и сохраняют общую динамику ряда, не создавая физически неправдоподобных скачков (рисунок 5).

Диагностика выбросов для сахарной свёклы. Для сахарной свёклы – культуры с единичными исторически аномально высокими значениями урожайности – подсчитано число self-masking-восстановленных значений (таблица 4), превышающих физически неправдоподобный порог 500 ц/га, отдельно по каждому из четырёх методов (36 наблюдений – объединение self-masking выборки по обоим вариантам пайплайна, без аугментации и с ней).

Ни один из четырёх методов не сгенерировал в self-masking ни одного физически неправдоподобного значения по сахарной свёкле – агрономическая маска и клиппинг постобработки эффективно предотвращают подобные артефакты во всех сравниваемых методах, а не только в VAE.

Пост-обработка и контроль качества. После импутации выполнено сглаживание резких скачков урожайности и проверка консистентности показателя Production для городских регионов с площадью посевов ниже 500 га (скорректировано от 10 до 68 строк в разных прогонах, максимальное единичное изменение – до 21,5 тыс. тонн в варианте с аугментацией). Пост-обработка агрономических ограничений обнулила 2356 невозможных значений в обоих вариантах; операция clip (lower = 0) применена 107 раз без аугментации и 59 раз с аугментацией – то есть аугментированная модель реже генерирует отрицательные значения, требующие принудительной обрезки. Итоговая валидация не выявила ни одного агрономического нарушения в обоих вариантах; отдельная проверка целостности не обнаружила подозрительных кросс-региональных совпадений (например, между регионами Ош и Ош City, которые в коде по ошибке ранее могли сравниваться как подстроки, – проверка на 15 показателях по обоим регионам, включая 31 совместный пропуск по сахарной свёкле, утечки не выявила). Финальная проверка региона Нарын (13 показателей культур, агрономически запрещённых в этом регионе) подтвердила нулевые значения по всем ячейкам в обоих вариантах пайплайна.

Обсуждение

MICE превосходит VAE на всех контрольных культурах. Результат раздела 7 подтверждает и уточняет вывод, сделанный ранее по данной линии исследования в статье [1]: MICE устойчиво точнее VAE на всех трёх контрольных культурах (хлопок, сахарная свёкла, пшеница) в обоих вариантах пайплайна. В отличие от предыдущей качественной оценки, здесь доступны точные значения RMSE: разрыв VAE–MICE варьируется от умеренного на пшенице (3,64 против 3,43 без аугментации, + 6 %) до существенного на сахарной свёкле (94,53 против 75,60, + 25 %) и хлопке (6,34 против 3,60, + 76 %).

Аугментация ухудшает, а не улучшает качество VAE. Наиболее значимое наблюдение этого прогона – аугментация обучающей выборки одновременно (а) ухудшает валидационный loss обучения VAE (0,0616 → 0,1449) и (б) ухудшает self-masking точность восстановления на всех трёх контрольных

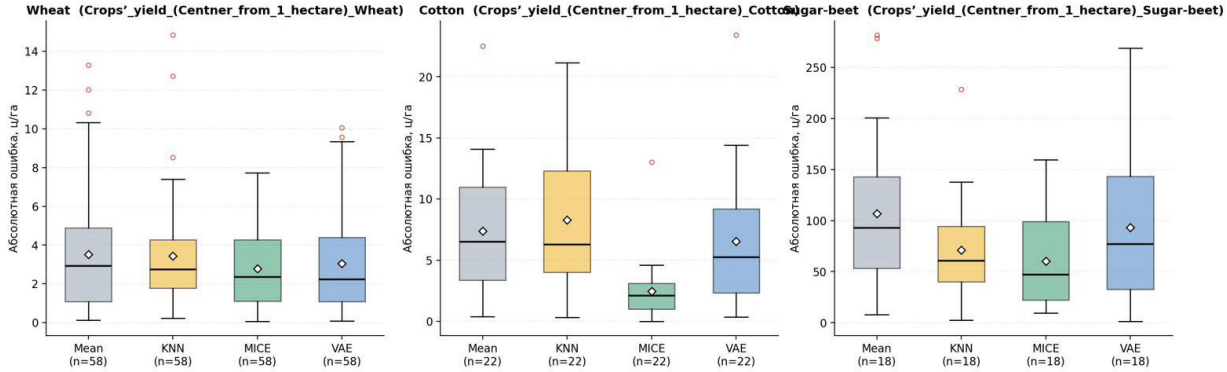


Рисунок 4 – Box-plot абсолютных ошибок self-masking восстановления по четырём методам (Mean, KNN, MICE, VAE) для трёх контрольных культур (вариант с аугментацией)



Рисунок 5 – Сравнение истинных и VAE-импутированных значений урожайности пшеницы по регионам Чуй, Ош и город Ош; для города Ош крестиками отмечены 5 фактически импутированных (реально пропущенных) значений

Таблица 4 – self-masking для сахарной свёклы

Метод	Значений > 500 ц/га	Всего наблюдений	Доля, %
Mean	0	36	0,0
KNN	0	36	0,0
MICE	0	36	0,0
VAE	0	36	0,0

культурах без исключения (на 11–34 % по RMSE). Это прямо противоречит рабочей гипотезе, заложенной в методологическое описание [1], согласно которой аугментация должна компенсировать малый объём данных и улучшать обобщающую способность модели. Наиболее вероятное объяснение – сама аугментация вносит дополнительный, специально спроектированный шум (гауссов шум, климатические сценарии, MixUp, bootstrap, маскирование), который расширяет обучающее распределение, но одновременно отдаляет его от распределения контрольных self-masking наблюдений, взятых из исходных 315 записей; в результате модель, обученная на аугментированном пуле, чуть хуже подгоняется под точные значения реальных, не аугментированных наблюдений. Это не означает, что аугментация бесполезна как приём регуляризации в общем случае, но означает, что в текущей конфигурации (масштабы шума, доля аугментированных примеров, отсутствие взвешивания вклада реальных и синтетических примеров в loss) она не даёт заявленного улучшения и должна быть переосмыслена как нейтральный или отрицательный результат либо перепроверена с иными гиперпараметрами аугментации перед повторным утверждением о её пользе [9, 10].

Практическая надёжность метода. Несмотря на отставание от MICE по точности, VAE в обоих вариантах пайплайна не генерирует агрономически невозможных значений, проходит все проверки на утечку данных и корректно обнуляет структурно недопустимые комбинации «регион – культура». Оценка неопределённости через MC Dropout остаётся содержательно работоспособной: пример с урожайностью хлопка в Баткене (1991) иллюстрирует, что для части реально пропущенных наблюдений модель даёт узкое, уверенное распределение прогнозов, тогда как для других – существенно более широкое (максимальное std в нормализованном пространстве – 0,82 без аугментации), что можно использовать как самостоятельный сигнал надёжности конкретных восстановленных значений независимо от итогового сравнения с MICE.

Заключение. Статья задокументировала фактический запуск пайплайна VAE-импутации на панели [2] в двух вариантах и объяснила его результаты: структуру признакового пространства (315 × 109), объём и характер пропусков (25,6 % сырых данных), кривые обучения ELBO, эффект аугментации на корреляционную структуру (среднее отклонение 0,094) и на loss (ухудшение с 0,0616 до 0,1449), оценку неопределённости MC Dropout на конкретном примере и self-masking валидацию на трёх культурах.

Два содержательных вывода требуют внимания при интерпретации результатов. Во-первых, MICE устойчиво превосходит VAE по точности восстановления на всех контрольных культурах в обоих вариантах пайплайна. Во-вторых, аугментация обучающей выборки в её нынешней конфигурации не улучшает, а ухудшает как валидационный loss, так и self-masking точность VAE – это требует либо пересмотра ожиданий от аугментации, либо дополнительной настройки её гиперпараметров перед повторной проверкой гипотезы о её пользе.

Тем не менее отставание VAE от MICE по точечной точности не отменяет его самостоятельных преимуществ, не воспроизводимых классическим методом в использованной здесь реализации:

Совместное восстановление всех признаков. VAE моделирует распределение сразу по 109 признакам одновременно, а не восстанавливает каждый показатель отдельной регрессией, как MICE – это позволяет учитывать нелинейные и многосторонние зависимости между климатом, экономикой и урожайностью, а не только парные связи.

Количественная оценка неопределённости. MC Dropout даёт для каждого восстановленного значения не только точечную оценку, но и доверительный интервал – IterativeImputer в использованной конфигурации выдаёт только точечный результат без меры уверенности.

Явный учёт агрономических ограничений. Взвешенная функция потерь ELBO встраивает агрономическую маску прямо в обучение модели, а не как отдельный шаг постобработки, и это работает: ни VAE, ни другие методы не сгенерировали ни одного физически неправдоподобного значения по сахарной свёкле.

Устойчивость к утечкам и структурным ограничениям. Модель корректно проходит проверки на утечку данных между регионами и на обнуление недопустимых комбинаций «регион – культура».

Иными словами, разрыв с MICE – это разрыв в точечной точности при малом объёме данных, а не в содержательных возможностях метода: там, где важны совместное моделирование признаков, оценка неопределённости или встроенные агрономические ограничения, VAE предлагает то, что MICE в данной реализации не даёт.

Поступила: 03.08.2026; рецензирована: 17.08.2026; принята: 19.08.2026.

Литература

1. Вариационный автоэнкодер для восстановления пропусков в панели агроэкономических данных Кыргызской Республики: архитектура, аугментация и валидация методом self-masking // Настоящий цикл исследований, статья в печати // Вестник КРСУ. 2026. Т. 26. № 8.
2. Формирование сбалансированной панели агроэкономических данных Кыргызской Республики: источники, структура и методы восстановления пропусков // Настоящий цикл исследований, статья в печати // Известия КГТУ им. И. Раззакова. 2026. № 3 (79).
3. *Картанова А.Д.* Создание базы верифицированных данных о сельхозпроизводителях: шаг к цифровой трансформации АПК / А.Д. Картанова, Ж.Р. Сарыпбекова, Ж.Б. Мамадалиева, З.А. Сыдыкова // Вестник Международного университета Кыргызстана. 2025. № 2-1 (59). С. 71–81. URL: https://www.elibrary.ru/item.asp?id=89032482 (дата обращения: 13.07.2026).
4. *Gneiting T., Raftery A.E.* Strictly Proper Scoring Rules, Prediction, and Estimation // Journal of the American Statistical Association. 2007. Vol. 102. № 477. P. 359–378.
5. *Gal Y., Ghahramani Z.* Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning // Proceedings of the 33rd ICML. 2016. P. 1050–1059.
6. *Kendall A., Gal Y.* What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? // Advances in Neural Information Processing Systems (NeurIPS). 2017. Vol. 30. P. 5574–5584.
7. *van Buuren S., Groothuis-Oudshoorn K*. Mice: Multivariate Imputation by Chained Equations in R // Journal of Statistical Software. 2011. Vol. 45. № 3. P. 1–67.
8. Простой метод импьютации данных победил // AI Hub. 2026. 11 янв. URL: https://aihubnews.ru/articles/prostoy-metod-impyutatsii-dannykh-pobedil (дата обращения: 13.07.2026).
9. *Makridakis S., Spiliotis E., Assimakopoulos V.* Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward // PLOS ONE. 2018. Vol. 13. № 3. e0194889.
10. Особенности работы с данными, характеризующими здоровье населения: заполнение пропусков в данных // Медицинский альманах. 2022. № 4. С. 50–61. URL: http://vestnik.mednet.ru/content/view/1139/30/lang,ru_ru.cp1251/ (дата обращения: 5.06.2026).