

РЕЦЕНЗИЯ № 1

на статью «ВОССТАНОВЛЕНИЕ ПРИПОВЕРХНОСТНОЙ ВЛАЖНОСТИ АТМОСФЕРЫ НАД ОКЕАНОМ ПО ДАННЫМ СОПУТСТВУЮЩИХ МЕТЕОРОЛОГИЧЕСКИХ ИЗМЕРЕНИЙ С ПРИМЕНЕНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ»

авторского коллектива: С. А. Вострикова, М. А. Криницкий, С. К. Гулев, М. П. Александрова

Данные метеорологических наблюдений играют критически важную роль в информационном обеспечении наук об атмосфере и климате. Качество метеорологической информации, полнота и точность данных являются важными факторами формирования адекватных представлений о процессах, протекающих в климатической системе Земли и ее отдельных компонентах. Наиболее изменчивой во времени и в пространстве составляющей атмосферного воздуха является водяной пар. Имеющиеся сведения о пространственном распределении влажности воздуха, особенно над океанами, не в полной мере удовлетворяют исследователей и практиков. В этой связи рецензируемая статья, в которой рассмотрен вопрос о восстановлении данных о приповерхностной влажности представляется важной и актуальной.

Для достижения цели авторы использовали несколько алгоритмов машинного обучения и сопоставили результаты восстановления данных, полученные с помощью рассмотренных алгоритмов. Авторами проделан огромный объем работы, поскольку подготовка входных данных, подбор гиперпараметров, обучение моделей, и последующая их валидация – это довольно рутинный процесс. На мой взгляд, формулировка моделей, их обучение и валидация выполнены вполне корректно. Анализ полученных результатов осуществлен в полной мере и свидетельствует о том, что алгоритмы машинного обучения представляют собой перспективный инструмент решения задач восстановления разнообразных данных о состоянии окружающей природной среды. Статья хорошо структурирована. Отдельные грамматические неточности не портят в целом положительного впечатления о представленном материале.

Общие замечания по статье:

1. Имеются ли какие-либо соображения на счет того, почему для разбиения по пространству использовалась сетка 2 на 2 градуса (*строка 171*)?
2. Нет ссылки на *рис. 3*.
3. Хочу заметить, что процедура «dropout» – коварная вещь. Ее применение не всегда приводит к желаемому результату, а вычислительное время может увеличиться в разы.

Редакционные замечания по тексту статьи:

1. *Строка 13*: «Влажность воздуха ...влияние на процессы переноса влаги...». Масло масляное. Влияет на самое себя. Требуется редакция.
2. *Строка 25*: «В качестве первого, наиболее простого подхода, мы решали задачу в формулировке...». Предлагаю заменить на «В статье задача решена в...».
3. *Строки 46, 47*: «**Влажность** воздуха над океаном в приповерхностном слое атмосферы является важным климатическим параметром, характеризующим потоки влаги и...». Характеризует самое себя. Требуется редакция.
4. *Строки 95, 110, 132, 136, 139, 142*: «...коллекцию данных...» заменить на «базу данных» или «набор данных». Словосочетание «коллекция данных» обычно используется в программировании: «Коллекция – структура данных (тип, класс, даже лучше сказать интерфейс), которая создана, чтобы содержать в себе некоторое количество объектов (в зависимости от языка и терминологии они должны быть одного типа или могут быть разных типов)».

5. *Строка 166*: «...характерна...» заменить на «...характеризуется...».
6. *Строки 223, 224*: «...суть которого... которая разделяет...». Требуется редакция.
7. *Строки 237, 238*: «Это условие...» заменить на «Этим условием...».
8. *Строка 298*: «...используя алгоритм оптимизации, в данной работе выбран Adam (ввиду его высокой...». Скобку надо поставить после запятой, а не перед словом «ввиду».

Резюме: принять без значительных правок, устранив редакционные неточности.

Подпись. Рецензент № 1. 17.03.2025.

От редакции: рецензия и файл статья с замечаниями были направлены авторскому коллективу.

Ответ рецензенту № 1 на Рецензию от 17.03.2025 на статью авторского коллектива: С. А. Вострикова, М. А. Криницкий, С. К. Гулев, М. П. Александрова «ВОССТАНОВЛЕНИЕ ПРИПОВЕРХНОСТНОЙ ВЛАЖНОСТИ АТМОСФЕРЫ НАД ОКЕАНОМ ПО ДАННЫМ СОПУТСТВУЮЩИХ МЕТЕОРОЛОГИЧЕСКИХ ИЗМЕРЕНИЙ С ПРИМЕНЕНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ».

Мы благодарим рецензента за время, уделенное нашей статье, и усилия, вложенные в оценку значимости наших результатов и их достоверности.

Далее мы приводим ответ на замечания рецензента по пунктам. Замечания рецензента приводятся полужирным шрифтом, исходный текст статьи приводится курсивом, наши комментарии и скорректированный текст приводится обычным шрифтом.

Рецензент: *1. Имеются ли какие-либо соображения насчет того, почему для разбиения по пространству использовалась сетка 2 на 2 градуса (строка 171)?*

Ответ: В отношении разбиения по пространству мы следовали зарекомендовавшему себя подходу регионализации в задачах, связанных с обработкой данных попутных наблюдений (например, ICOADS), применяемому, в частности, в (Alexandrova et al., 2021). В других работах (Alexandrova, Gulev and Belyaev, 2018) применяется более укрупненное пространственное деление с использованием 5-градусных квадратов. В нашем исследовании мы отталкивались от баланса между наполненностью выделяемых регионов данными и специфичностью таких регионов.

Рецензент: *2. Нет ссылки на рис. 3.*

Ответ: Мы дополнили текст статьи, прокомментировав содержимое рис. 3 в Разделе 2.2 «Методы» (L200-201). В частности, мы добавили следующий текст:

Распределение количества наблюдений в квадратах $2 \times 2^\circ$ в наборе данных DISO3 представлено на рисунке 3.

Рецензент: *3. Хочу заметить, что процедура «dropout» – коварная вещь. Ее применение не всегда приводит к желаемому результату, а вычислительное время может увеличиться в разы.*

Ответ: Действительно, процедура прореживания не всегда дает ожидаемый результат в случае нейросетевых моделей и может привести как к снижению показателей качества, так и к существенному повышению общего времени обучения моделей. В нашем исследовании не применялся метод прореживания (dropout) ни в одной из моделей. Этот подход упомянут лишь в качестве одного из доступных методов регуляризации, которые позволяют гибко настраивать искусственные нейронные сети и балансировать их выразительную и обобщающую способности.

Рецензент: *Строка 13. «Влажность воздуха ...влияние на процессы переноса влаги...». Масло масляное. Влияет на самое себя. Требуется редакция.*

Ответ: Мы благодарим рецензента за это замечание. Мы исправили текст статьи так, чтобы он читался более ясно.

Исходный текст в аннотации:

Влажность воздуха в приповерхностном слое атмосферы над океаном является ключевым климатическим параметром, оказывающим значительное влияние на процессы переноса влаги и тепла между океаном и атмосферой, а также на динамику атмосферных процессов в целом.

Исправленный текст:

Влажность воздуха в приповерхностном слое атмосферы над океаном является ключевым климатическим параметром, определяющим потоки влаги и тепла на границе океана с атмосферой, и оказывающим значительное влияние на динамику атмосферных процессов в целом.

Рецензент: *Строка 25: «В качестве первого, наиболее простого подхода, мы решали задачу в формулировке ...». Предлагаю заменить на «В статье задача решена в...».*

Ответ: Действительно, формулировку в аннотации можно упростить без искажения смысла.

Исходный текст в аннотации:

В качестве первого, наиболее простого подхода, мы решали задачу в формулировке аппроксимации относительной влажности...

Скорректированный текст:

В настоящей статье решена задача в формулировке аппроксимации относительной влажности...

Рецензент: *Строки 46, 47 «Влажность воздуха над океаном в приповерхностном слое атмосферы является важным климатическим параметром, характеризующим потоки влаги и тепла на границе океана с атмосферой и влияющим на динамику атмосферы». Характеризует самое себя. Требуется редакция.*

Ответ: Спасибо за замечание. Действительно, формулировка была неточной. Мы отредактировали строки 46 и 47, чтобы текст читался более ясно.

Исходный текст в разделе «Введение»:

Влажность воздуха над океаном в приповерхностном слое атмосферы является важным климатическим параметром, характеризующим потоки влаги и тепла на границе океана с атмосферой и влияющим на динамику атмосферы.

Скорректированный текст:

Влажность воздуха над океаном в приповерхностном слое атмосферы – важный климатический параметр, определяющий процессы переноса влаги и тепла на границе океана с атмосферой и влияющий на динамику атмосферных процессов.

Рецензент: *Строки 95, 110, 132, 136, 139, 142: «... коллекцию данных..» заменить на «базу данных» или «набор данных». Словосочетание «коллекция данных» обычно используется в программировании: «Коллекция — структура данных (тип, класс, даже лучше сказать интерфейс), которая создана, чтобы содержать в себе некоторое количество объектов (в зависимости от языка и терминологии они должны быть одного типа или могут быть разных типов)».*

Ответ: Вы правы, термин «коллекция данных» чаще используется в контексте программирования. Мы заменили словосочетание «коллекция данных» на «базу данных» и «набор данных» во всех указанных местах.

Исходный текст в разделе «Введение»:

На первом этапе мы сформировали коллекцию данных DISO3...

Эта коллекция стала основой для обучения и оценки качества алгоритмов.

Скорректированный текст:

На первом этапе мы сформировали набор данных DISO3...

Эта база данных стала основой для обучения и оценки качества алгоритмов.

Исходный текст в разделе 2.1 «Данные»:

В исследовании использовалась коллекция данных DISO3 (Dataset of In-Situ Observations over the Ocean).

Рис. 1 – Пространственная плотность наблюдений и измерений коллекции данных ICOADS (Freeman et al., 2017) за период с 1950 по 2017 гг., в логарифмической шкале.

Рис. 2 – Общее годовое количество наблюдений в коллекции данных DISO3.

Таблица 1 – Состав переменных коллекции данных DISO3.

Коллекция DISO3 содержит пропуски, поэтому перед применением методов машинного обучения данные были стандартизированы...

Скорректированный текст:

В исследовании использовался набор данных DISO3 (Dataset of In-Situ Observations over the Ocean).

Рис. 1 – Пространственная плотность наблюдений и измерений базе данных ICOADS (Freeman et al., 2017) за период с 1950 по 2017 гг., в логарифмической шкале.

Рис. 2 – Общее годовое количество наблюдений в наборе данных DISO3.

Таблица 1 – Состав переменных набора данных DISO3.

Набор данных DISO3 содержит пропуски, поэтому перед применением методов машинного обучения данные были стандартизированы...

Исходный текст в разделе 2.2 «Методы»:

Рис. 3 — Распределение количества наблюдений в квадратах $2 \times 2^\circ$ в коллекции данных DISO3.

Скорректированный текст:

Рис. 3 — Распределение количества наблюдений в квадратах $2 \times 2^\circ$ в наборе данных DISO3.

Рецензент: *Строка 166: «...характерна ...» заменить на «...характеризуется ...».*

Ответ: Мы согласны с замечанием, замена «...характерна...» на «...характеризуется...» в строке 166 улучшает читабельность и делает предложение более понятным.

Исходный текст в разделе 2.2 «Методы»:

«...относительная влажность приповерхностного слоя атмосферы характерна существенными сезонными вариациями».

Скорректированный текст:

«...относительная влажность приповерхностного слоя атмосферы характеризуется существенными сезонными вариациями».

Рецензент: *Строки 223, 224: «...суть которого... которая разделяет...». Требуется редакция.*

Ответ: Благодарим за внимательность к деталям. Вы совершенно правы, в строках 223 и 224 наблюдалась несогласованность рода. Мы исправили это предложение.

Исходный текст в разделе 2.2 «Методы»:

Дерево решений — это алгоритм машинного обучения, суть которого заключается в построении древовидной структуры, которая разделяет данные на подмножества...

Скорректированный текст:

Дерево решений – это алгоритм машинного обучения, который строит иерархическую структуру (дерево) для последовательного разделения данных на подмножества.

Рецензент: *Строки 237, 238: «Это условие...» заменить на «Этим условием...».*

Ответ: Благодарим за замечание. В строках 237 и 238 мы заменили «Это условие...» на «Этим условием...», чтобы обеспечить грамматическую связность.

Исходный текст в разделе 2.2 «Методы»:

Это условие может быть, например, достижение максимальной глубины дерева...

Скорректированный текст:

Этим условием может быть, например, достижение максимальной глубины дерева...

Рецензент: *Строка 298: «...используя алгоритм оптимизации, в данной работе выбран Adam (ввиду его высокой...». Скобку надо поставить после запятой, а не перед словом «ввиду».*

Ответ: Вы совершенно правы, скобка была поставлена некорректно. Мы перенесли закрывающую скобку после запятой.

Исходный текст в разделе 2.2 «Методы»:

Затем, используя алгоритм оптимизации, в данной работе выбран Adam (ввиду его высокой эффективности и устойчивости)...

Скорректированный текст:

Затем, используя алгоритм оптимизации, (в данной работе выбран Adam ввиду его высокой эффективности и устойчивости)...

СПИСОК ЛИТЕРАТУРЫ:

1. Alexandrova M. CLOUDINESS OVER THE OCEANS AT SUBARCTIC LATITUDES AS A VISIBLE PART OF ATMOSPHERIC MOISTURE TRANSPORT // Russian Journal of Earth Sciences. 2021. № 1 (21). С. 4.
2. Aleksandrova M., Gulev S. K., Belyaev K. Probability Distribution for the Visually Observed Fractional Cloud Cover over the Ocean // Journal of Climate. 2018. № 8 (31). С. 3207–3232.

С уважением, авторский коллектив. 23.04.2025.

От редакции: ответ и доработанная версия статьи были направлены редакцией рецензенту.

Подтверждение Рецензента № 1 на публикацию:

Авторы в полной мере учли мои замечания, комментарии и пожелания.

Резюме: принять в представленном виде.

Подпись. Рецензент № 1. 28.04.2025.