

Обзор поточечных и комплексных мер качества в исследованиях атмосферы и океана

Резвов В.Ю. (1,2), Криницкий М.А. (1,2)

(1) Московский физико-технический институт

(2) Институт океанологии им. П.П. Ширшова РАН



Резвов Вадим Юрьевич,
н.с. – зам.зав.лабораторией
машинного обучения в науках о Земле,
МФТИ

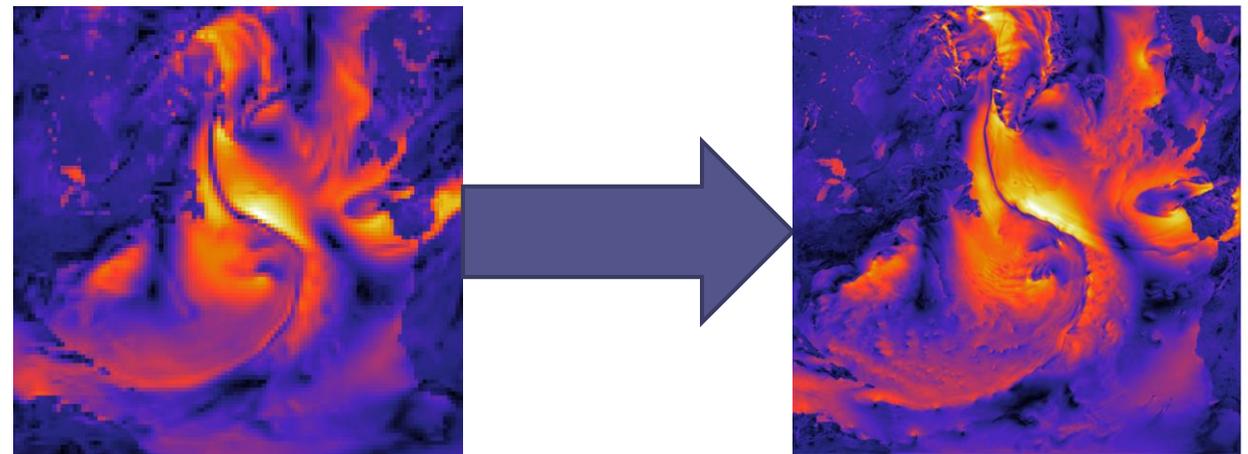
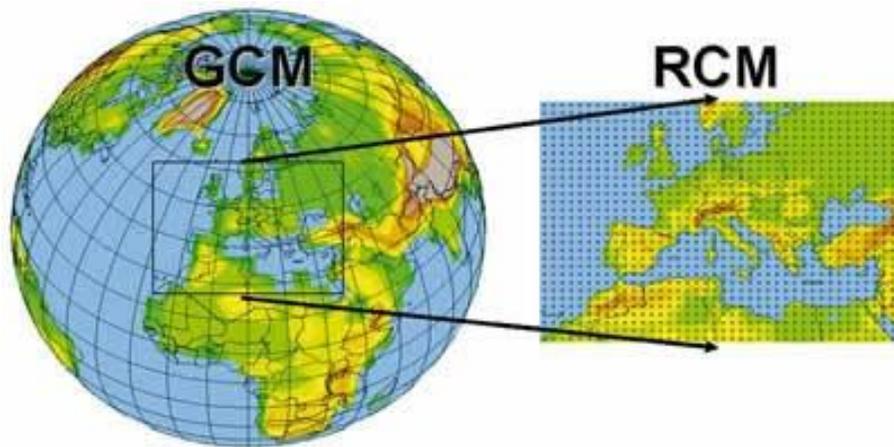
Москва, 20 июня 2024

Актуальность исследования

- Изменение климата и его последствия оказывают значительное влияние на общество – необходимость точного прогнозирования погоды и климата – **важность правильной оценки того, как модель воспроизводит реальные процессы.**
- Необходима разработка **метрик, или мер, качества**, представляющих результат моделей в легко интерпретируемом виде.
- **Метрики качества** – количественные показатели, упрощающие сложные, многомерные наборы данных до кратких числовых или категориальных представлений.
- **Метрики обеспечивают сравнение между собой климатических моделей** и оценку их пригодности для различных задач, что обеспечивает выбор наиболее релевантных данных для конкретных ситуаций и прикладных применений моделей.
- **Цель данного исследования** – обзор наиболее часто применяемых в задачах атмосферы и океана метрик качества.

Актуальность исследования

- Метрики качества важны в применении к сеточным метеорологическим данным
- Повышение актуальности применения метрик в задачах масштабирования данных (получение **мелкомасштабной** информации о физических величинах **из имеющихся** данных моделирования с **низким пространственным разрешением**) о погоде и климате: динамического и статистического.



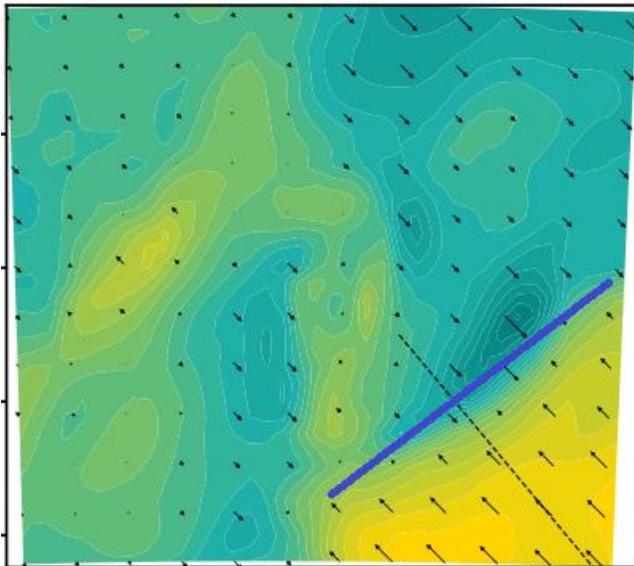
Классификация метрик

- Поточечные метрики основаны на поточечных ошибках между предсказанными моделью y и наблюдаемыми значениями x переменных:
- смещение (B) $\frac{1}{N} \sum_{i=1}^N (y_i - x_i)$
- средняя абсолютная ошибка (MAE) $\frac{1}{N} \sum_{i=1}^N |y_i - x_i|$
- средний квадрат ошибки (MSE) $\frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2$
- пиковое отношение сигнала к шуму (PSNR) $10 \lg \left(\frac{[\max_i \{x_i\}]^2}{MSE} \right)$
- среднеквадратическая ошибка (RMSE) \sqrt{MSE} – стандартный статистический показатель для оценки качества модели в метеорологии и климате
- Проблемы RMSE:
 - придает ошибкам с более высокими абсолютными значениями больший вес в сравнении с MAE;
 - ошибки модели должны соответствовать нормальному распределению.

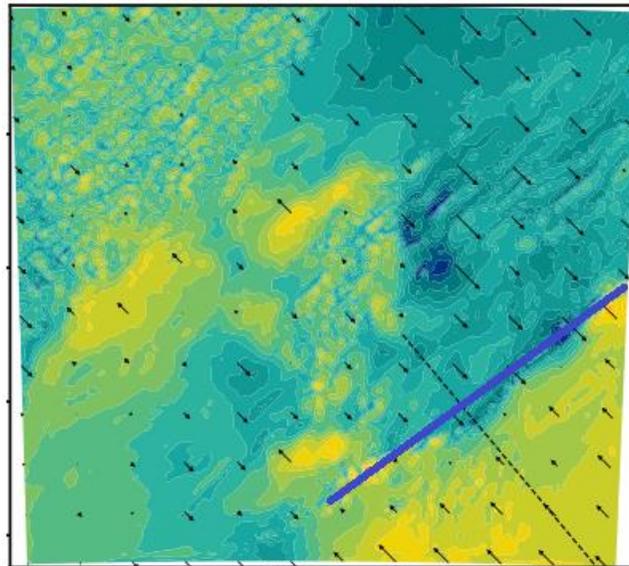
Общие проблемы поточечных метрик

- Для пространственных полей метеорологических переменных характерна **когерентная пространственная структура**.
- Поточечные методы оценки качества **не учитывают** присущую таким полям внутреннюю пространственную корреляцию.
- Результат поточечных метрик **сложно интерпретируем** в физическом смысле.

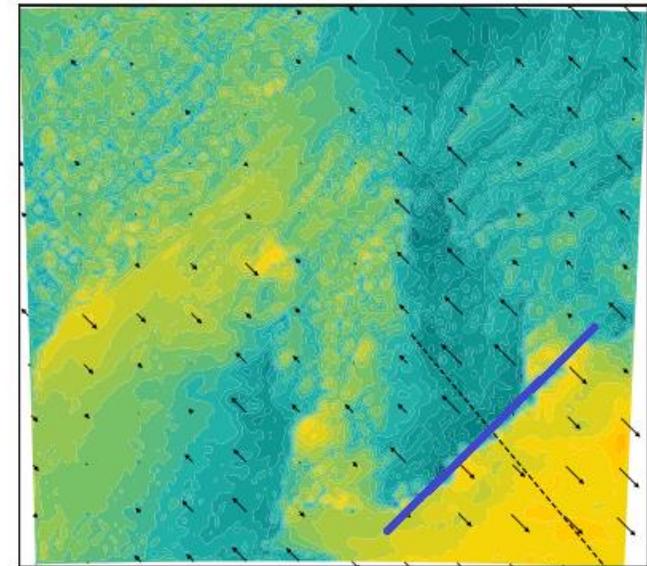
Пример: положение атмосферного фронта



Грубое разрешение



Результат масштабирования



Высокое разрешение

Комплексные метрики

- Количественная оценка качества модели требует **комплексных показателей**, фокусирующихся на различных аспектах качества прогноза погоды.
- Направления разработки комплексных метрик:
 - метрики, не зависящие от модели ошибок;
 - феноменологические метрики;
 - метрики, основанные на бинарных событиях, например, наличие/отсутствие осадков;
 - метрики, связанные с оценкой воспроизведения экстремальных событий и другие.

Комплексные метрики

- Различные модельные реализации полей могут иметь сходные показатели MSE, а их пространственные когерентные структуры – сильно отличаться.
- **Индекс структурной схожести (SSIM)** - приблизительная аппроксимация визуально воспринимаемой разницы в пространственной структуре между двумя изображениями. SSIM с помощью скользящего окна оценивает различия в яркости, контрасте и структуре изображений (или сеточных данных).
- Преимущества: локальная метрика SSIM может учитывать ковариационную структуру ошибок в целом.
- Недостатки: произвольность выбора масштаба скользящего окна; определение через средние значения, дисперсии и корреляции недостаточно хорошо моделирует восприятие зрительной системой человека.

Комплексные метрики

- Идея для комплексной метрики – комбинация таких метрик – **эффективность Клинга-Гупты (KGE)**:

$$KGE = 1 - \sqrt{(R-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$$

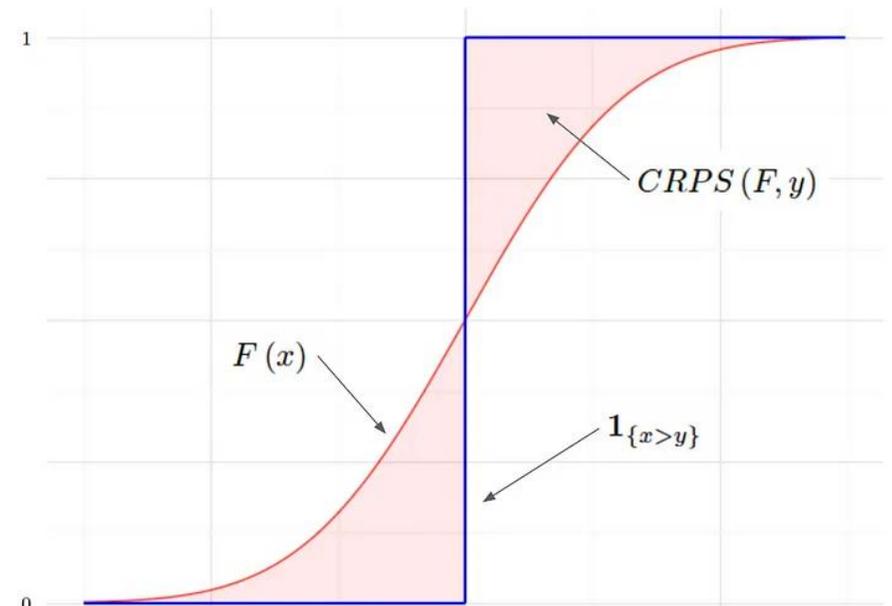
$$\alpha = \frac{\sigma_y}{\sigma_x} \quad \beta = \frac{\bar{y}}{\bar{x}} \quad R - \text{коэффициент корреляции}$$

- KGE может принимать значения от $-\infty$ до 1, при этом значение 1 соответствует идеальному согласию между моделированием и фактическими значениями.
- Недостаток: нет акцента на пространственной структуре, сохранение недостатков поточечных метрик.

Комплексные метрики

- **Оценка непрерывной ранжированной вероятности (CRPS)** – расширение метрики MAE на случай вероятностных или ансамблевых прогнозов метеорологических переменных
- CRPS показывает интегрированную по всему диапазону ошибку распределения прогнозируемой величины относительно распределения детерминированного значения.

CRPS чувствительна ко всему непрерывному диапазону интересующей переменной и не требует введения ряда predetermined классов, от которых могут зависеть результаты.

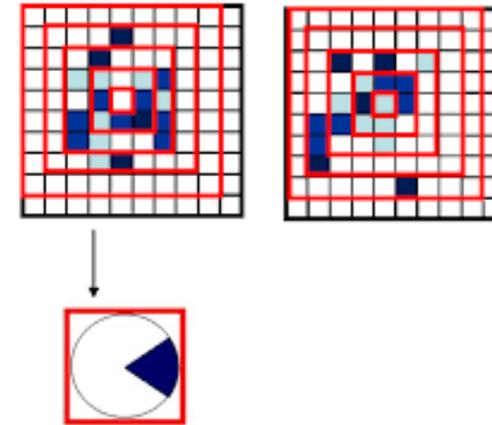


Комплексные метрики

Показатель навыков дробей (FSS) разработан изначально для оценки моделирования осадков

Рассматривается доля ячеек пространственной сетки в скользящем окне, занятых явлением, в поле прогноза y и в поле наблюдений x для всех положений скользящего окна:

$$FSS = 1 - \frac{\sum_{i=1}^N (P_{y,i} - P_{x,i})^2}{\sum_{i=1}^N P_{x,i}^2 + \sum_{i=1}^N P_{y,i}^2}$$



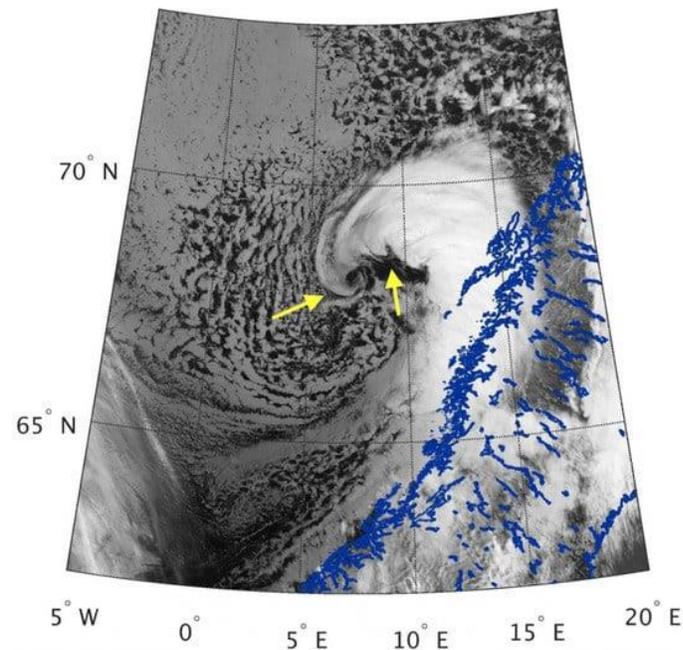
Преимущества: способность определить пространственный масштаб, на котором модель можно считать имеющей приемлемое качество; FSS связана с пространственным смещением осадков в прогнозе, что отчасти имитирует визуальную оценку качества прогноза осадков.

Комплексные феноменологические метрики

Феноменологические метрики основаны на идентификации «объектов» и сравнении их атрибутов (например, интенсивности, положения центроидов) по данным моделирования и фактическим данным.

Гибкость по отношению к атрибутам, присущим разным «объектам».

Пример: оценка воспроизведения моделью когерентных мезомасштабных структур.



Комплексные феноменологические метрики

- Для экстремальных явлений необходимо правильно оценивать качество прогнозирования метеорологических переменных с помощью специальных метрик.
- Самый простой подход: экстремальное событие как бинарная категория с точки зрения его (не)наступления. Качество оценивается с использованием традиционных категориальных оценок для бинарных событий. Такие метрики имеют тенденцию к неинформативным пределам 0 или 1.
- Пример более сложной метрики – **смещение 98 перцентиля**
- Редкость событий может привести к небольшому или нулевому количеству событий при разделении событий по классам, и при небольших размерах выборки могут оказывать значительное влияние выбросы, искажающие проверку прогнозов экстремальных явлений.
- Метрики, связанные с оценкой воспроизведения экстремальных событий, подвергают всю модель «дилемме прогнозиста», т.е. занижают оценку качества модели, оценивая только редкие по времени и пространству события

Заключение

- Несмотря на универсальность применения к сеточным данным, **поточечные методы обладают рядом существенных недостатков**, включая неспособность выделения качества воспроизведения внутренней пространственной корреляции полей и **чувствительность к сдвигам неоднородностей**.
- Недостатки классических метрик позволяют выделить некоторые общие тенденции в развитии метрик качества:
- Разработка методов, учитывающих пространственную структуру и неоднородности метеорологических полей, а также статистику данных.
- Разработка методов вероятностной оценки качества ансамблевых прогнозов.
- Преодоление проблем, присущих метрикам экстремальных явлений.
- Разработка алгоритмов идентификации важных «объектов» в пространственных полях для создания объектно-ориентированных метрик.
- Направление дальнейших исследований - отход от повсеместного применения классических метрик и переход к более комплексным метрикам, специфичным для различных задач атмосферы и океана.