Модуль 2 · Реанализ и многомерные климатические данные

Реанализ и данные наблюдений: сравнение и оценка ошибок

Ph.D., старший преподаватель Мадина Турсумбаева

Кафедра метеорологии и гидрологии

Курс: Современные цифровые технологии в метеорологии

Цель лекции

Научить сравнивать данные реанализа ERA5 с наблюдениями метеорологической станции: совмещать ряды по времени и месту, рассчитывать метрики ошибок и объяснять причины расхождений.

Лабораторная работа «Сравнение данных реанализа ERA5 с наблюдениями» — пример для Алматы.

Результаты обучения

После лекции студент будет:

Знать

Метрики: смещение (bias), MAE, RMSE, коэффициент корреляции

Причины расхождений: высота, представительность, суточный ход

Уметь

Перевести время из UTC в местное

Совместить ряд ERA5 и станции

Посчитать ошибки и построить диаграмму рассеяния

Понимать

ERA5 — оценка, а не наблюдение

Результат зависит от периода, места и срока

Зачем сравнивать реанализ с наблюдениями

ERA5

Средняя по ячейке около 31 км, высота — средняя по ячейке, время — UTC, часовой шаг.

Станция

Точка с реальной высотой и локальными условиями; время — местное; возможны пропуски и ошибки прибора.

Вопрос: насколько ERA5 пригоден там, где станций нет? Ответ даёт сравнение на станциях, где они есть.

Шаг 1: время UTC и местное время

```
import pandas as pd
# ERA5: время в UTC → местное время Алматы
s.index = s.index.tz_localize("UTC").tz_convert("Asia/Almaty")
# станция: время уже местное, добавляем пояс
obs.index = obs.index.tz_localize("Asia/Almaty")
```

Часовой пояс Алматы

С 1 марта 2024 г. вся территория Казахстана живёт по UTC+5. До этого Алматы был в UTC+6: для ранних лет проверяйте пояс данных.

Почему это важно

Сдвиг на 1 час смещает суточный ход и занижает или завышает ошибку в одни часы суток.

Шаг 2: точка сетки и высота

```
# ближайший узел сетки ERA5 к станции Алматы, °C
era = ds["t2m"].sel(latitude=43.25, longitude=76.95, method="nearest") - 273.15
s = era.to_series()                    # время UTC
```

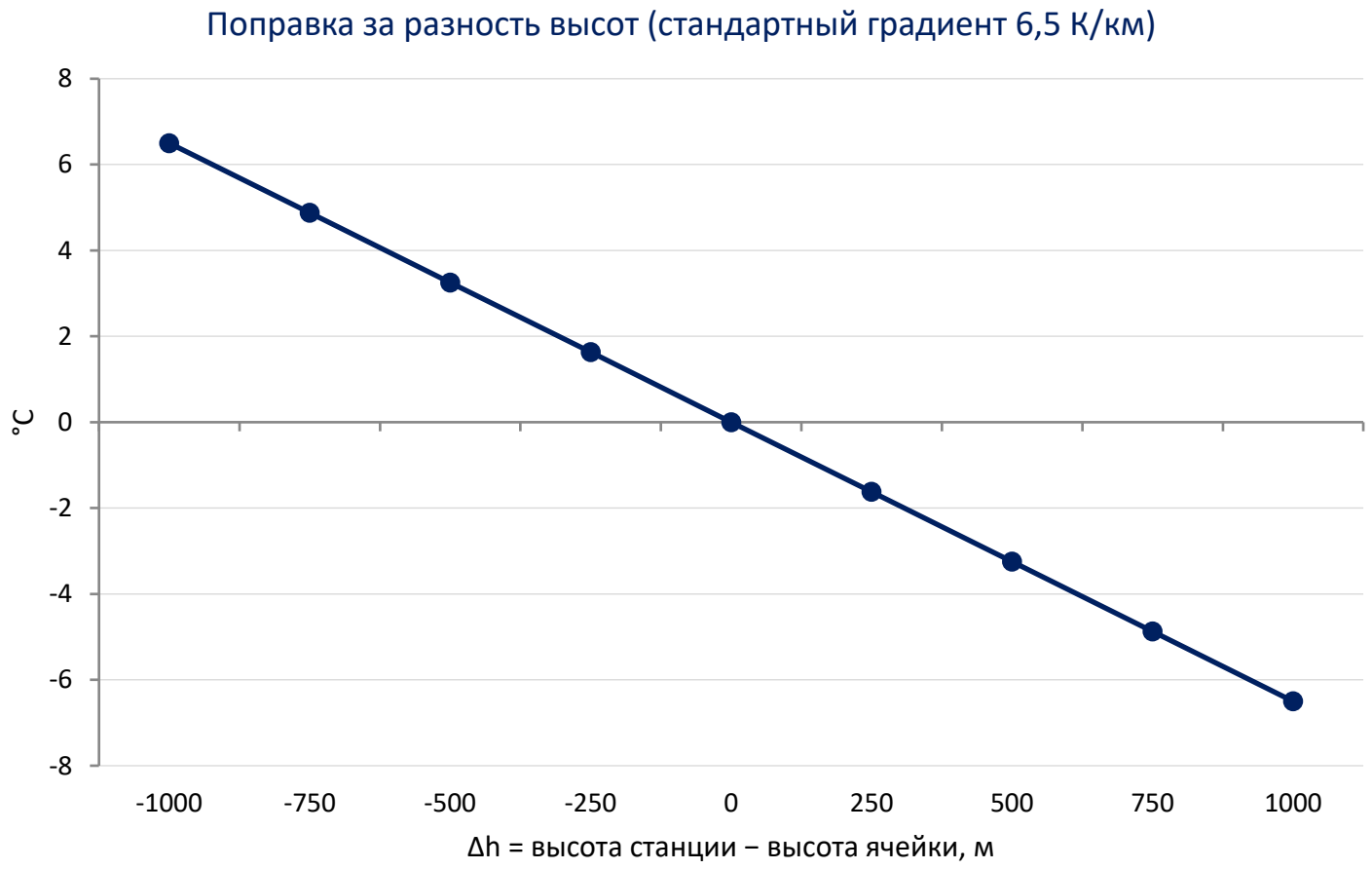
Поправка на высоту

Если высота станции и средняя высота ячейки различаются, температура отличается. При стандартном градиенте 6,5 К/км разность высот 100 м даёт примерно 0,65 °C.

Представительность

Ячейка 31 км объединяет город, предгорья и долины. Станция — одна точка. Различия — не ошибка модели, а разный масштаб.

Поправка на высоту: график



Как читать

Станция выше ячейки (Δh > 0) холоднее ERA5.

Станция ниже ячейки (Δh < 0) теплее ERA5.

Граница применимости

Градиент — стандартный. Реальный может отличаться, особенно ночью и в котловинах.

График — расчёт по формуле, а не данные ERA5 или станции.

Результаты ERA5 для Алматы: данные ещё не получены

Метрика	Значение	Как получить
bias, °C	—	mean(ERA5 − станция) по общим часам
MAE, °C	—	mean(|ERA5 − станция|)
RMSE, °C	—	sqrt(mean((ERA5 − станция)²))
r	—	corr(ERA5, станция)

Что нужно для заполнения

Загрузить ERA5 для точки станции по CDS (лекция 8) и файл наблюдений за тот же период.

Значения в таблице — только после расчёта на ваших данных. До этого выводы о качестве ERA5 не делаются.

Шаг 3: совмещение рядов

```
import numpy as np
obs = pd.read_csv("almaty_obs.csv", parse_dates=["time"]).set_index("time")["T"]
obs.index = obs.index.tz_localize("Asia/Almaty")
pair = pd.concat([s, obs], axis=1, keys=["era5", "obs"]).dropna()
print(len(pair), "общих часов")
```

Что делает concat и dropna

Соединяет два ряда по времени: остаются только совпавшие сроки, где есть значения у обоих рядов. Число общих часов всегда проверяйте.

Шаг 4: метрики ошибок

```
err = pair["era5"] - pair["obs"]
bias = err.mean()                          # смещение: ERA5 − станция
mae = err.abs().mean()                     # средняя абсолютная ошибка
rmse = np.sqrt((err ** 2).mean())          # среднеквадратичная ошибка
r = pair["era5"].corr(pair["obs"])         # корреляция
```

bias

Знак показывает, теплее ли ERA5.

MAE

Типичная величина ошибки без учёта знака.

RMSE

Сильнее штрафует крупные ошибки.

r

Согласованность изменений, не величину ошибки.

Шаг 5: диаграмма рассеяния

```
import matplotlib.pyplot as plt
lim = [pair.min().min(), pair.max().max()]
plt.scatter(pair["obs"], pair["era5"], s=6, alpha=0.4)
plt.plot(lim, lim, "k--", label="y = x")
plt.xlabel("Станция, °C"); plt.ylabel("ERA5, °C"); plt.legend()
```

- Точки выше линии y = x: ERA5 теплее станции; ниже — холоднее.
- Разброс вокруг линии — случайная ошибка; наклон облака, отличный от 1, — систематическая.

Шаг 6: ошибки по месяцам и часам

```
# смещение по месяцам: сезонная структура ошибки
err.groupby(err.index.month).mean().plot(marker="o")
# смещение по часам суток: суточный ход ошибки
err.groupby(err.index.hour).mean().plot(marker="o")
```

Что искать

Смещение, зависящее от сезона или часа, указывает на физическую причину: снежный покров, ночные инверсии, суточный ход.

Вывод

Среднее смещение около нуля не означает, что ошибок нет: разные знаки в разные сезоны взаимно уничтожаются.

Причины расхождений

Причина	Проявление	Что проверить
Высота станции и ячейки	Постоянное смещение	Поправка 6,5 К/км
Городская застройка и котловина	Смещение, зависящее от часа	Расположение станции
Ночные инверсии	Ошибки в ночные часы	Суточный ход ошибки
Ошибки станции: прибор, пропуски	Выбросы, скачки	Контроль качества (лекция 4)
Смена систем наблюдений в ERA5	Изменение ошибки во времени	Ограничения (лекция 9)

Google Earth Engine: как войти

1 **Аккаунт Google**
Подойдёт любой: личный Gmail или университетский.

2 **Откройте редактор**
В браузере: code.earthengine.google.com — и войдите в аккаунт.

3 **Зарегистрируйте проект**
При первом входе откроется регистрация: создайте новый проект Google Cloud или выберите существующий.

4 **Выберите некоммерческое использование**
Укажите цель Noncommercial и тип проекта, заполните короткую анкету. Уровень Community — без платёжных данных.

5 **Запустите скрипт**
Откроется Code Editor. Проверьте, что вверху справа выбран ваш проект, вставьте код и нажмите Run.

Бесплатно для учёбы

Некоммерческий проект не требует оплаты.

Лимит уровня Community — 150 EECU-часов в месяц на проект; он рассчитан на студентов.

Преподаватель может добавить студентов в общий учебный проект.

Тот же вход из Python

```
import ee
ee.Authenticate()
ee.Initialize(project="my-project")
```

my-project — идентификатор вашего проекта Google Cloud.

Источник: Google Earth Engine Guides — Earth Engine access; Noncommercial tiers (developers.google.com/earth-engine/guides).

GEE Code Editor: что где находится



Схема интерфейса Code Editor (code.earthengine.google.com), не снимок экрана. Код в окне — условный пример.

1 Проект и аккаунт
Вверху справа. Проверьте, что выбран ваш проект.

2 Scripts · Docs · Assets
Ваши скрипты, справка по функциям, загруженные данные.

3 Редактор кода и Run
Сюда вставляют скрипт. Run запускает его.

4 Inspector · Console · Tasks
Значение в точке карты; вывод print и графики; экспорт.

5 Карта
Результат: слои, добавленные через Map.addLayer.

GEE для быстрой проверки (только температура)

```
// В GEE: ECMWF/ERA5/DAILY, band mean_2m_air_temperature (K)
// Сравнивать с станцией можно только по суточным средним
```

```
# на стороне Python: станция — суточные средние
obs_daily = obs.resample("1D").mean()
day = pd.concat([era.resample("1D").mean(), obs_daily], axis=1, keys=["era5", "obs"]).dropna()
```

Важно

GEE удобен для быстрой проверки температуры. Почасовое сравнение — в CDS и Python: суточные агрегаты GEE не дают часового хода.

Заключение

1

Время

UTC → Asia/Almaty

2

Точка

ближайший узел и высота

3

Совмещение

concat и dropna

4

Метрики

bias, MAE, RMSE, r

Главный вывод

Чтобы сравнить ERA5 с наблюдениями, ряды совмещают по времени (UTC и местное время) и по месту (узел сетки и высота). Смещение, MAE, RMSE и r оценивают вместе, по месяцам и часам. ERA5 — оценка, а не наблюдение, и результат зависит от периода, места и срока.

Вопросы для самопроверки

1. Почему сравнивать почасовые значения без перевода времени неверно?
2. Почему корреляция r высокая, а bias может быть ненулевым?
3. Что означает точка выше линии y = x на диаграмме рассеяния?
4. Почему нулевой bias не гарантирует точности ERA5 в каждый сезон?

Семинар 10 и следующий модуль

Семинар 10 · ERA5 и данные станции

Совмещение, метрики и диаграмма рассеяния для Алматы.

Лабораторная работа «Сравнение данных реанализа ERA5 с наблюдениями».

Далее: Модуль 2 · Лекция 11

Пространственный анализ ERA5: cartopy и MetPy.

Лекции 1–10 завершают первый набор курса.

Литература

- Hersbach, H., et al. (2020). The ERA5 global reanalysis. QJRMS, 146, 1999–2049. https://doi.org/10.1002/qj.3803
- Willmott, C. J. (1981). On the validation of models. Physical Geography, 2, 184–194.
- Wilks, D. S. (2019). Statistical Methods in the Atmospheric Sciences (4th ed.). Elsevier.
- Interfax (19 января 2024). Kazakhstan switching to single time zone from March 1.
- Gorelick, N., et al. (2017). Google Earth Engine. Remote Sensing of Environment, 202, 18–27.
- Raoult, B., Hersbach, H., et al. ERA5: The ECMWF Reanalysis Data. ECMWF, 2018.