

Двадцать третья международная конференция  
"СОВРЕМЕННЫЕ ПРОБЛЕМЫ ДИСТАНЦИОННОГО  
ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА"

Машинное обучение для коррекции остатков  
энергобалансовой модели с целью повышения  
точности оценок эвапотранспирации  
агроэкосистем

ФГБНУ «Агрофизический научно-  
исследовательский институт»  
Доброхотов А.В., Козырева Л.В, Мухина Д.П.  
[adobrokhotov@agophys.ru](mailto:adobrokhotov@agophys.ru)

10 - 14 ноября 2025 г., Москва

# Энергетический и водный баланс

- Энергетический баланс:

$$R_n = H + LE + G$$

$R_n$  (net radiation) — радиационный баланс. Доступная энергия (разница между приходящей и уходящей радиацией).

$H$  (sensible heat flux) — поток явного тепла. Энергия, которая нагревает воздух у поверхности.

$LE$  (latent heat flux) — поток скрытого тепла. Энергия, расходуемая на испарение воды.

$G$  (soil heat flux) — поток тепла в почву. Энергия, идущая на нагрев почвы.



$$ET = LE / \lambda$$

$ET$  (evapotranspiration) — суммарное испарение (испарение с поверхности почвы + транспирация)

$\lambda$  — удельная теплота парообразования ( $\sim 2.45 \times 10^6$  Дж/кг при температуре около  $20^{\circ}\text{C}$ )

- Водный баланс

$$\Delta S = P + IRR + C - ET - R - D$$

$\Delta S$  — изменение влагозапасов в почвенном профиле (зоне корней).

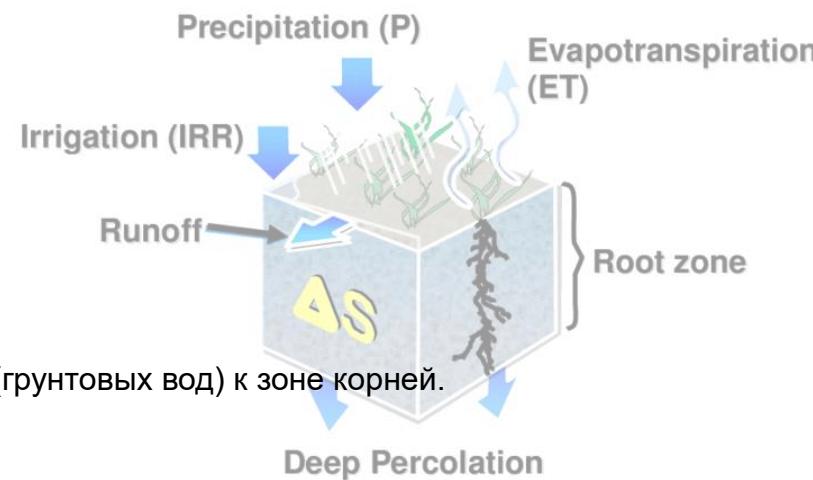
$P$  (precipitation) — атмосферные осадки (дождь, снег).

$IRR$  (irrigation) — орошение, поливная вода.

$C$  (capillary rise) — капиллярный подъем воды из более глубоких горизонтов (грунтовых вод) к зоне корней.

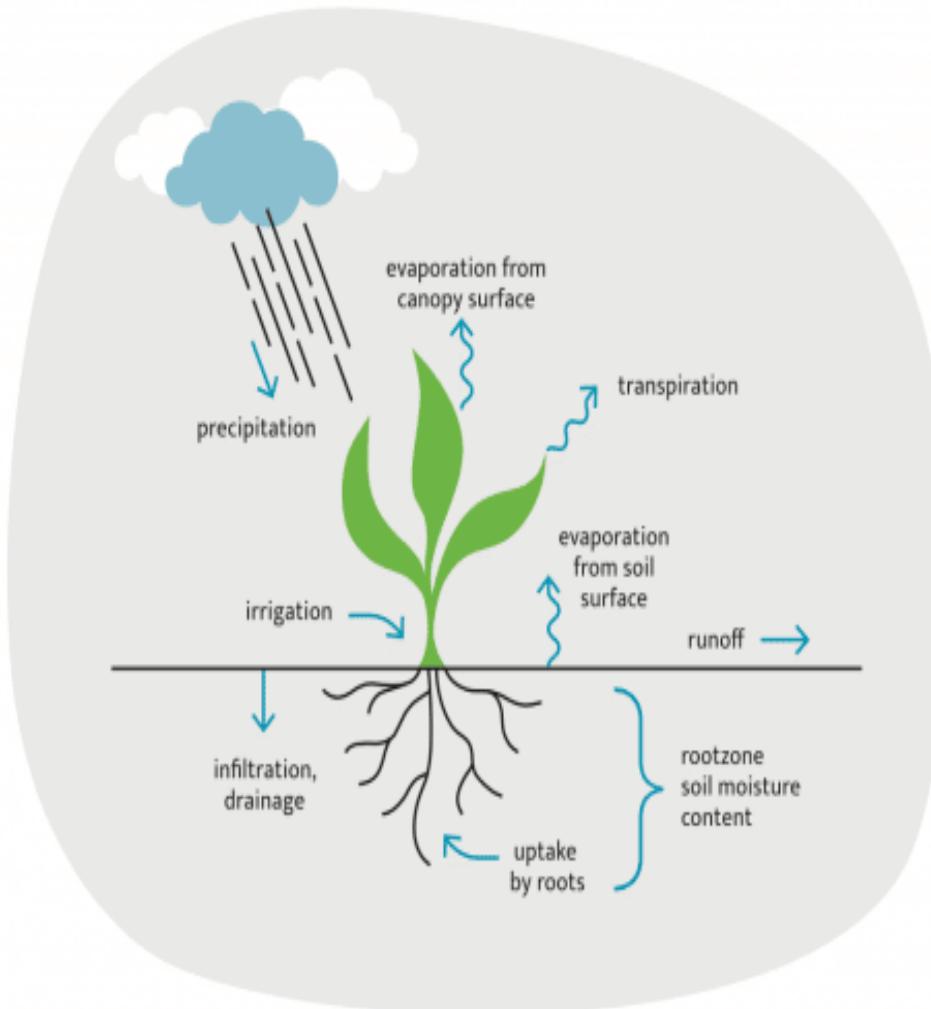
$R$  (runoff) — поверхностный сток.

$D$  (deep percolation) — глубокое просачивание (вода, которая уходит ниже зоны корней).

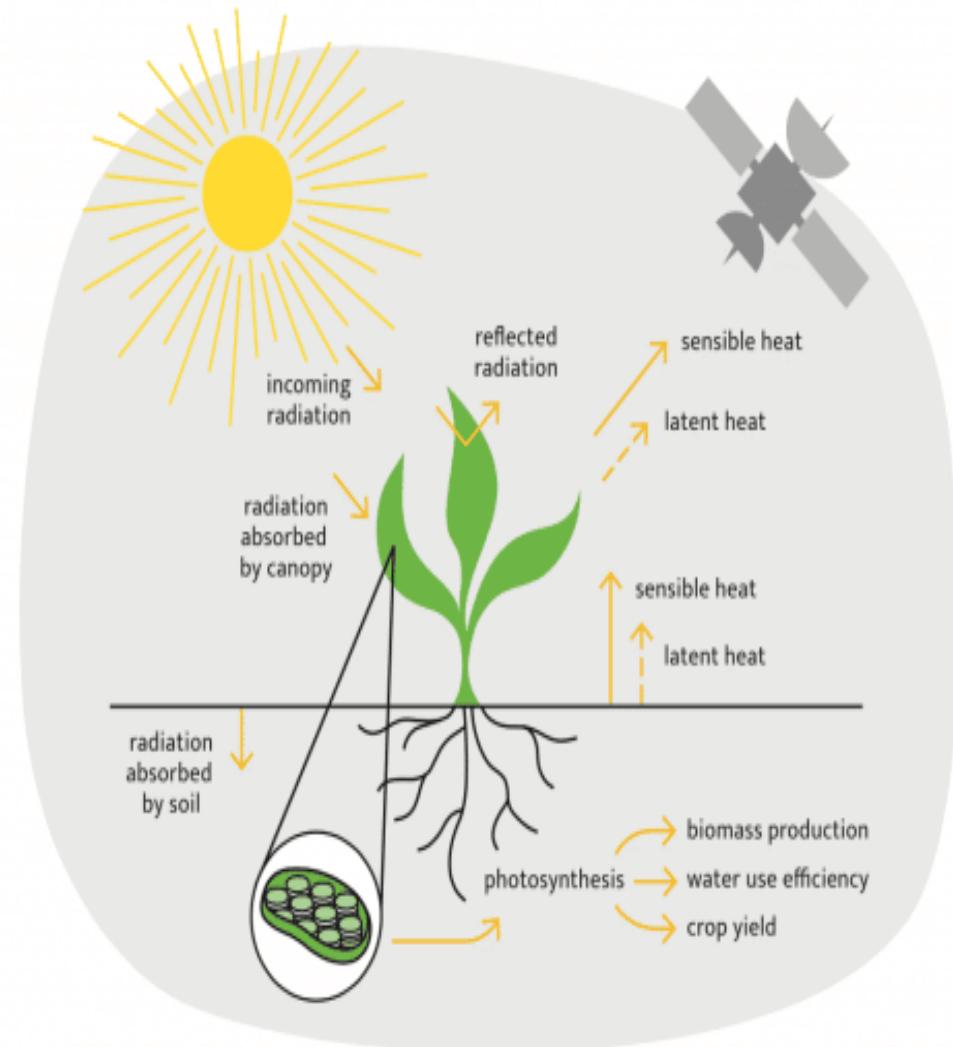


# Энергетический и водный баланс

## Water balance

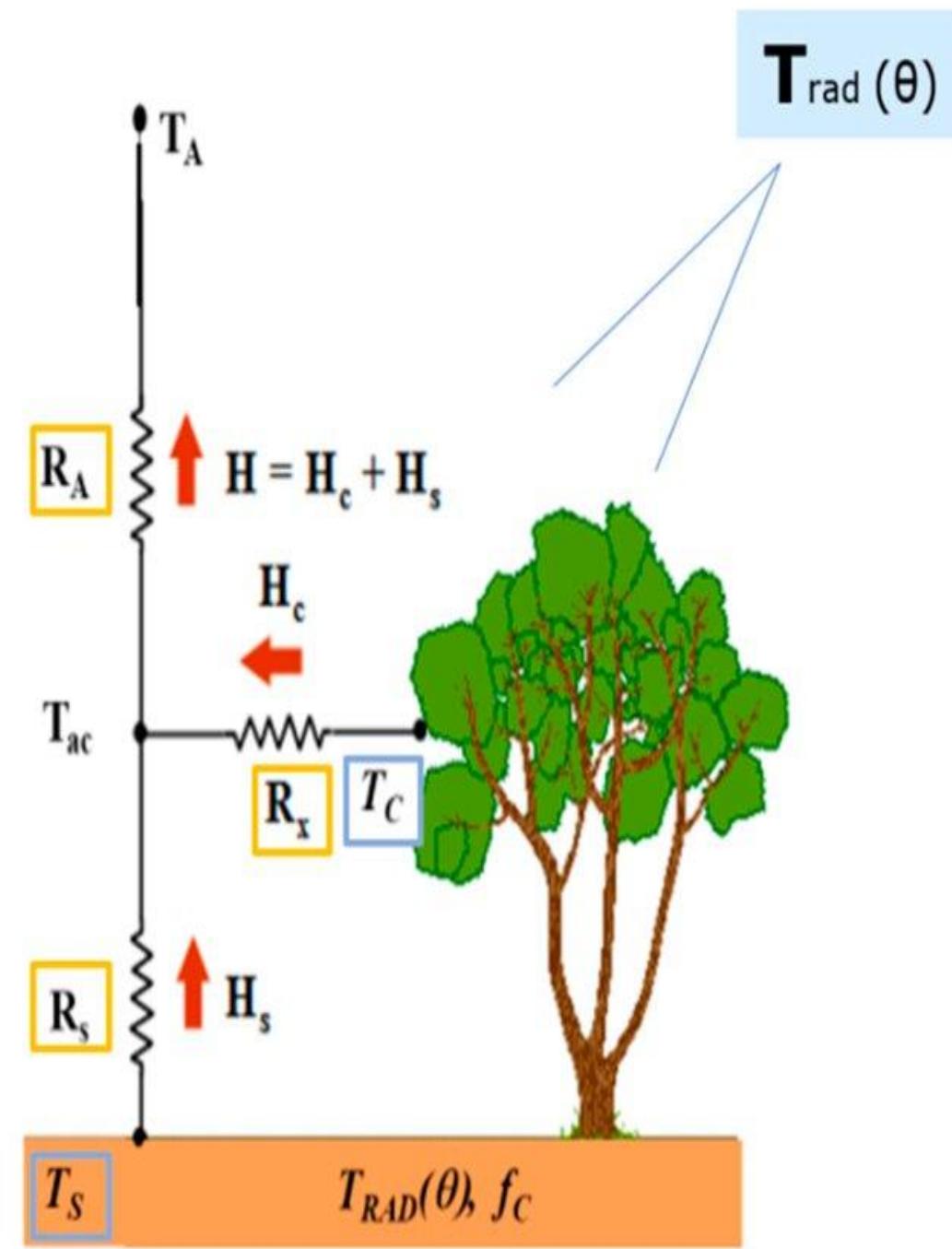


## Energy balance



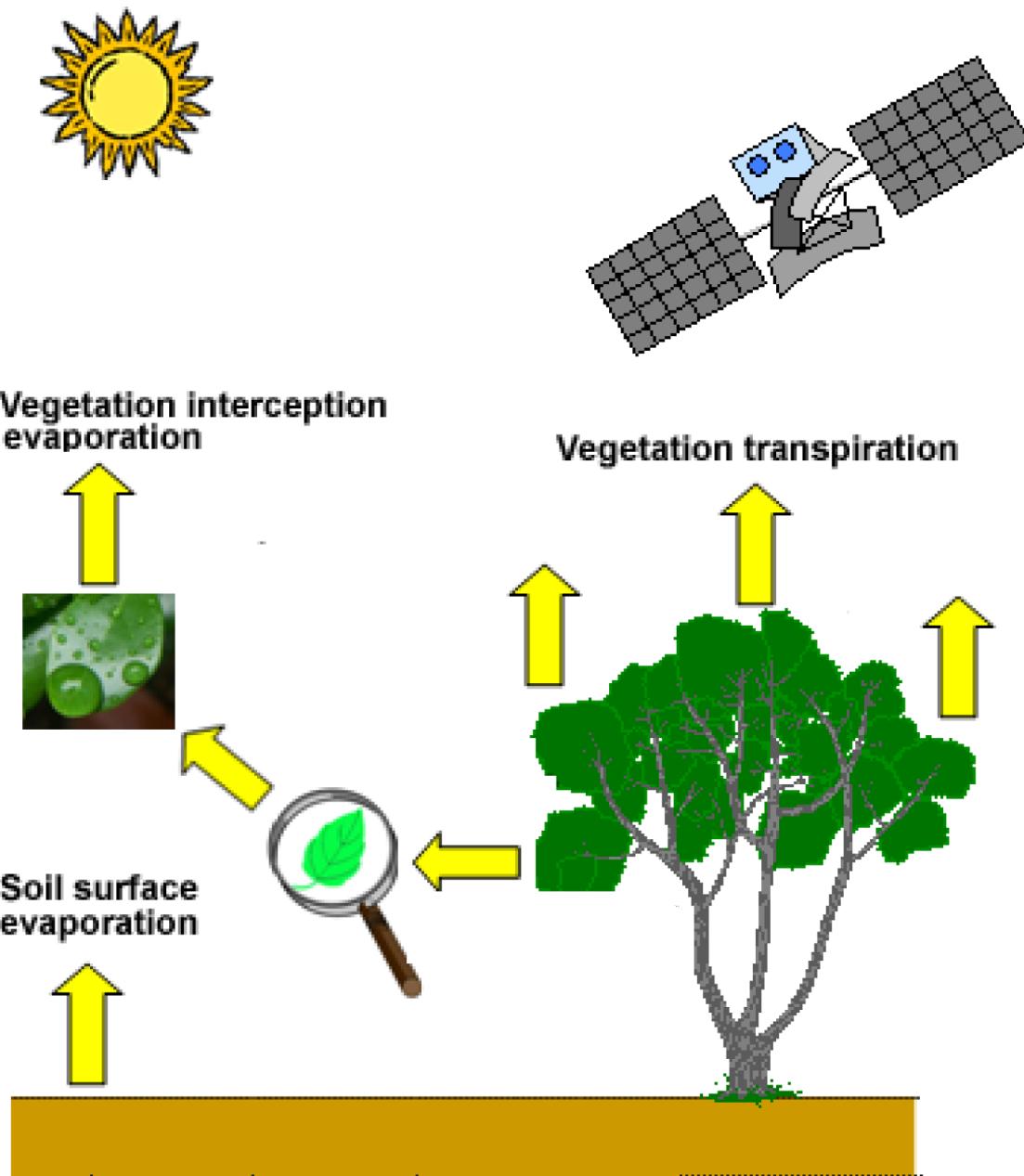
# Модель TSEB

- **TSEB (Two-Source Energy Balance)** – это двухуровневая энергобалансовая модель для оценки скрытого и явного потоков тепла над почвой (LEs) и растительностью (LEc) на основе спутниковых и метеорологических данных.
- **Входные данные:** 1) спутниковые снимки в тепловом инфракрасном диапазоне (для оценки температуры поверхности) и в видимом/ближнем ИК-диапазоне (для расчета LAI/NDVI); 2) метеорологические данные: скорость ветра, температура воздуха, влажность воздуха, приходящая коротковолновая и длинноволновая радиация
- **Разделение температуры:** Радиационная температура, измеренная спутником ( $T_{rad}$ ), разделяется на компоненты – температуру растений ( $T_c$ ), температуру почвы ( $T_s$ ), температуру воздуха в растительном покрове ( $T_a$ )
- **Аэродинамический подход:**  $R_A$  – аэродинамическое сопротивление (сопротивление потоку от воздуха в пологе к атмосфере);  $R_s$  – сопротивление почвы (сопротивление потоку от поверхности почвы к воздуху в растительном пологе);  $R_x$  – сопротивление листового пограничного слоя (сопротивление потоку от поверхности листа к воздуху в растительном пологе)
- + коррекция влияния стратификации на потоки с помощью масштаба Монина-Обухова и универсальных функций



# Модель PT-JPL

- **PT-JPL (Priestley-Taylor Jet Propulsion Laboratory)** – это упрощенная энергобалансовая модель для оценки скрытого потока тепла при испарении влаги с поверхности почвы (LEs), транспирации (LEc), а также влаги, перехваченной листьями (роса, осадки) (LEi) на основе спутниковых и метеорологических данных.
- **Входные данные:** 1) спутниковые снимки спутниковые снимки в тепловом инфракрасном диапазоне (для оценки температуры поверхности) и в видимом/ближнем ИК-диапазоне (для расчета LAI/NDVI); 2) метеорологические данные: скорость ветра, температура воздуха, влажность воздуха, приходящая коротковолновая и длинноволновая радиация
- **Факторы снижения потенциальных потоков:** максимально возможный скрытый поток тепла, расчетанные по формуле Пристли-Тейлора снижается в зависимости от лимитирующих испарение факторов:  $f_{wet}$  (коэффициент влажности поверхности),  $f_g$  (коэффициент проективного покрытия растительности),  $f_T$  (коэффициент температуры растительности),  $f_{sm}$  (коэффициент дефицита почвенной влаги)
- нет коррекции влияния стратификации



# ML алгоритм оптимизации

**Градиентный бустинг** — это техника машинного обучения (ML) для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений. Обучение ансамбля проводится последовательно. На каждой итерации вычисляются отклонения предсказаний уже обученного ансамбля на обучающей выборке. Следующая модель, которая будет добавлена в ансамбль будет предсказывать эти отклонения. Таким образом, добавив предсказания нового дерева к предсказаниям обученного ансамбля мы можем уменьшить среднее отклонение модели, которое является таргетом оптимизационной задачи.

В нашем случае мы будем строить регрессию между остатками (ошибки, разница измерениями и оценками) и энергобалансовой модели (TSEB или PT-JPL) и факторами окружающей среды (климат, почва, ДЗЗ).

В основе ML-оптимизированной энергобалансовой модели лежит физически обоснованная модель и коррекция ее ошибок + возможность разделения потоков на основе физической модели.



$f_0$

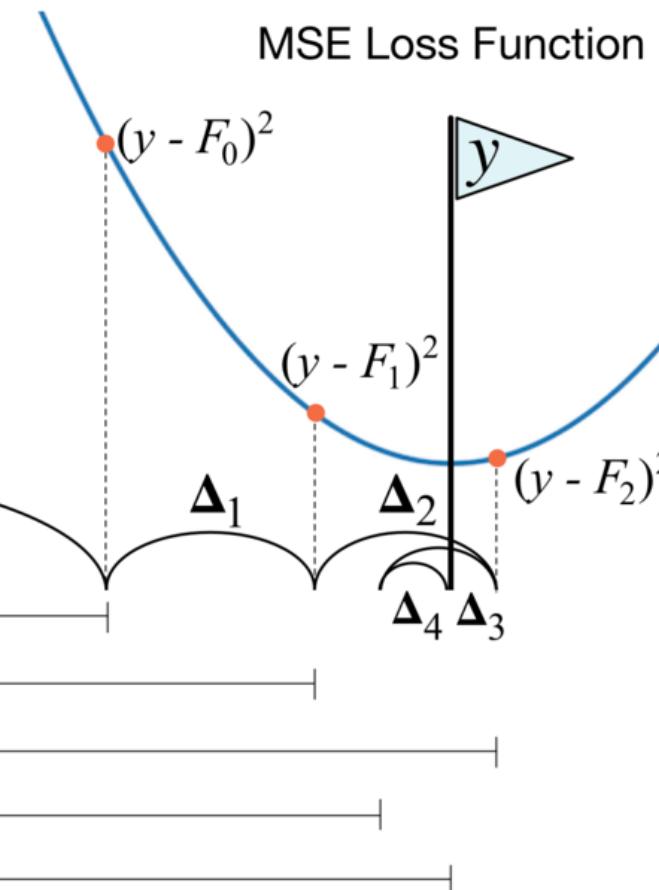
$F_0$

$F_1$

$F_2$

$F_3$

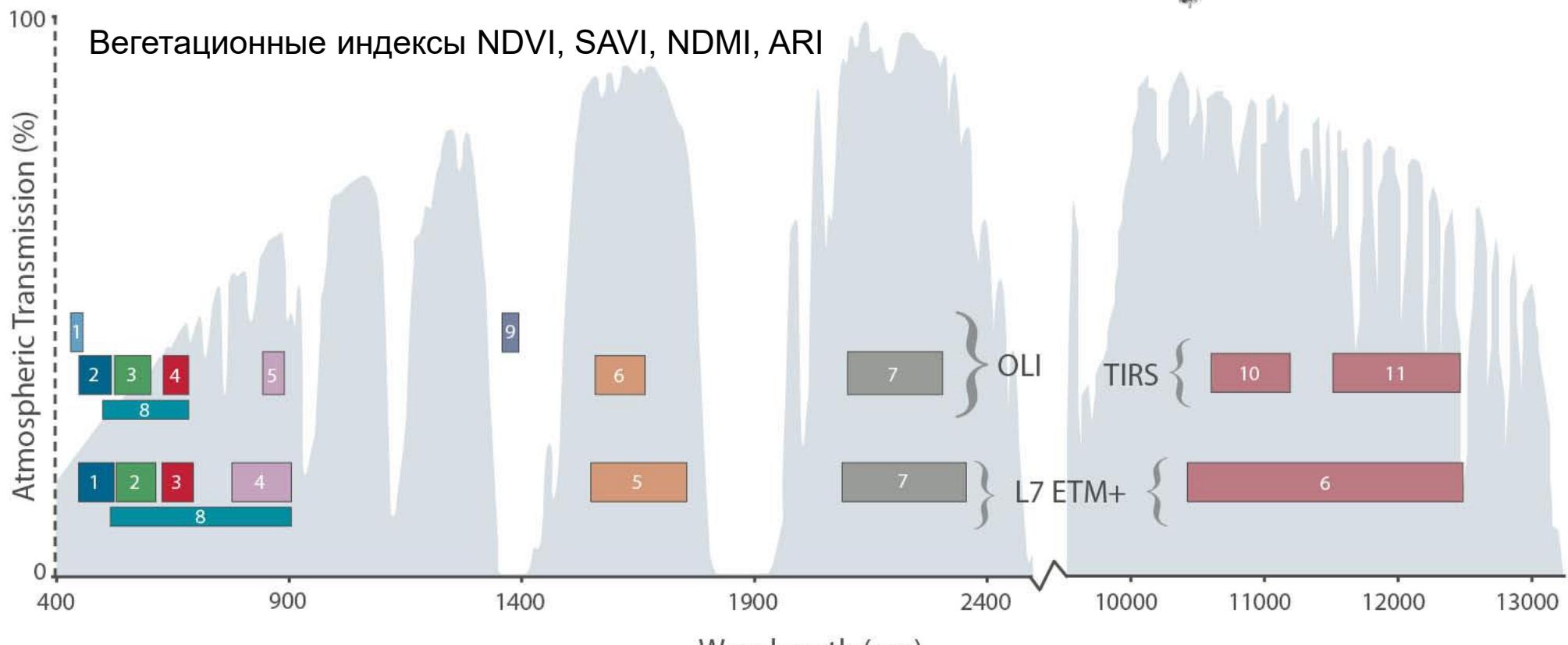
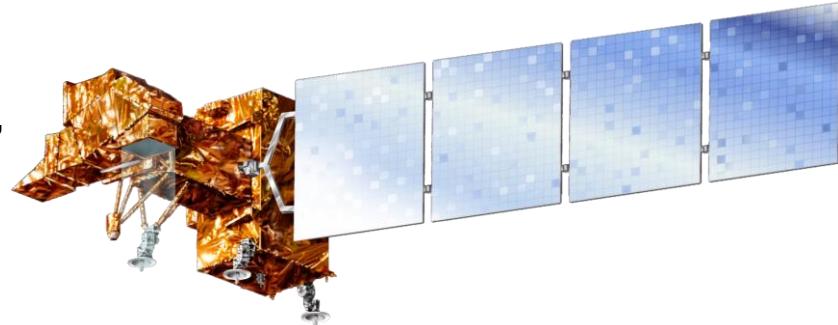
$F_4$



# Данные дистанционного зондирования (ДЗЗ)

Landsat, пространственное разрешение 30-100 метров.

Фильтрация снимков по наличию облачности и влияния теней, неопределенности оценки температуры поверхности, влияния аэрозолей. Всего отобрано **4880 снимков** для сельскохозяйственных полей с измерениями (метод турбулентных пульсаций) LE в момент пролета спутника.



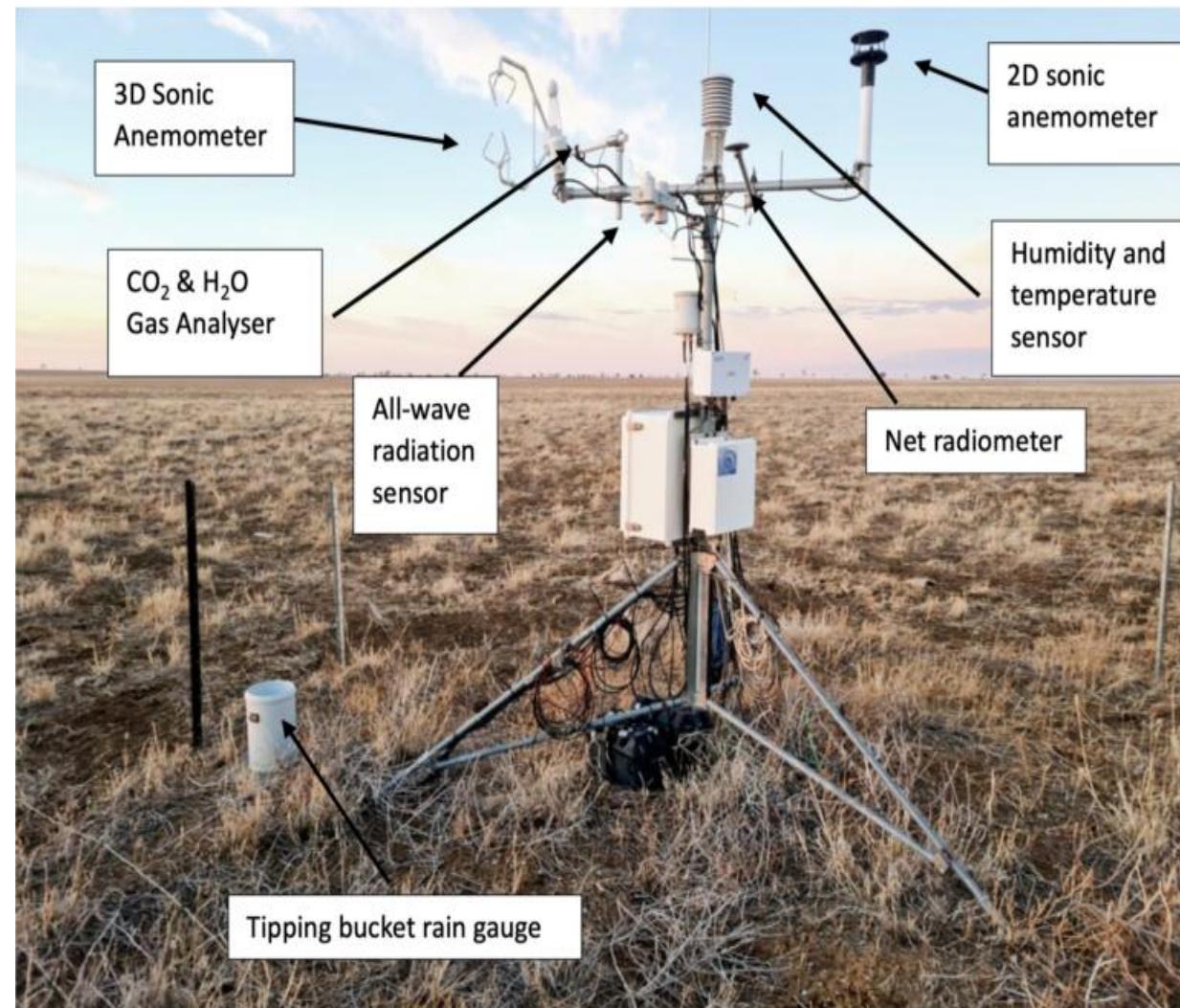
# Данные измерений LE

Данные измерений скрытого потока тепла (LE) методом турбулентных пульсаций с 137 станций, расположенных над сельскохозяйственными полями по всему миру.

**Метод турбулентных пульсаций** (eddy covariance) позволяет измерять LE, основываясь на корреляции пульсаций вертикальной скорости ветра ( $w'$ ) и пульсаций удельной влажности воздуха ( $q'$ ). Установленная на мачте система быстрореагирующих датчиков (3D-анемометр и анализатор влажности) с высокой частотой (например, 10 Гц) фиксирует эти микрофлуктуации, а математический расчёт ковариации между ними даёт прямое и непрерывное измерение потока парообразной влаги.

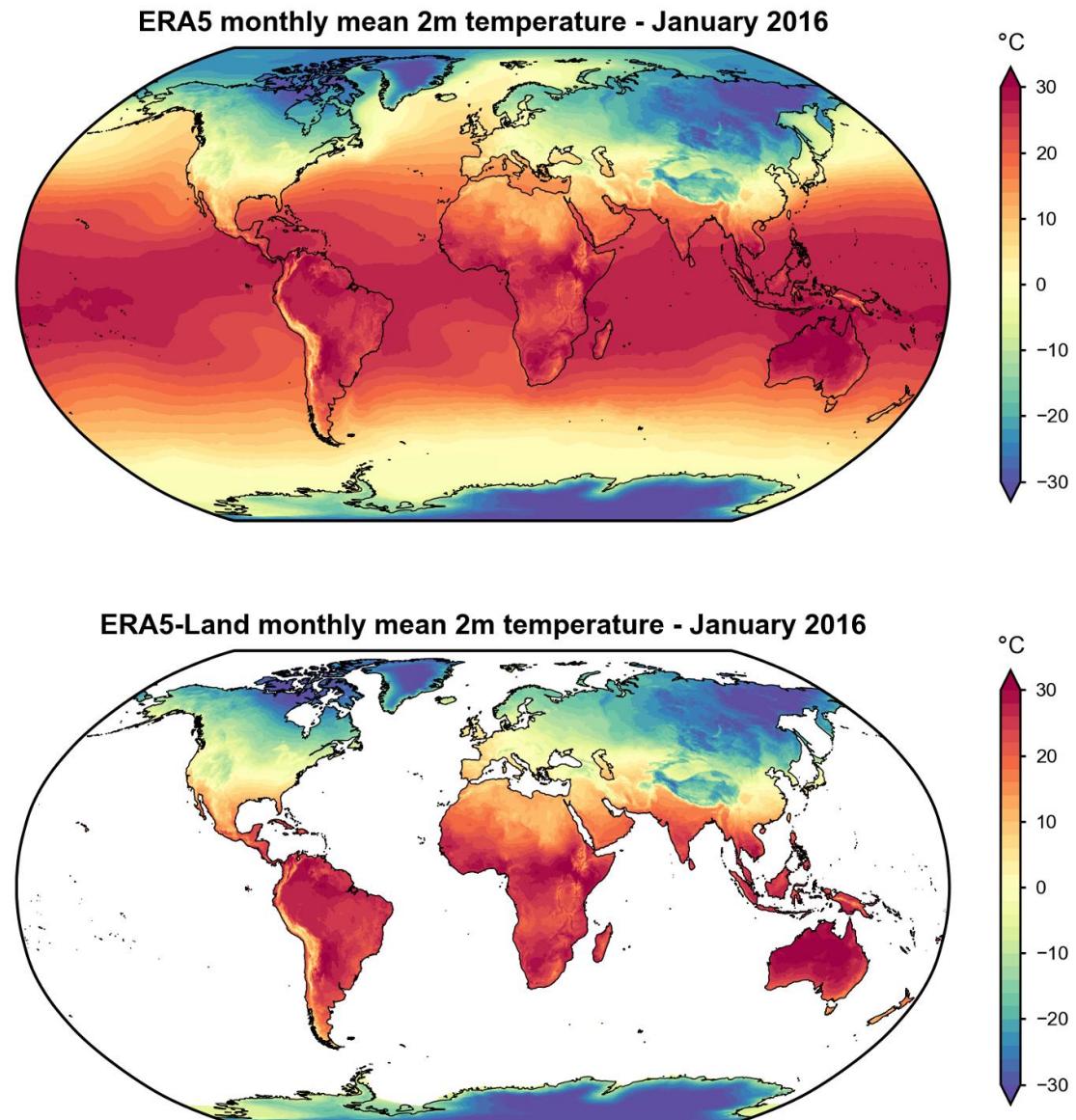
## Фильтрация данных:

- 1) по флагу качества (QA) при препроцессинге;
- 2) по замыканию энергетического баланса  $(R_n + G) — (LE + H) < 200 \text{ Вт}/\text{м}^2$ .



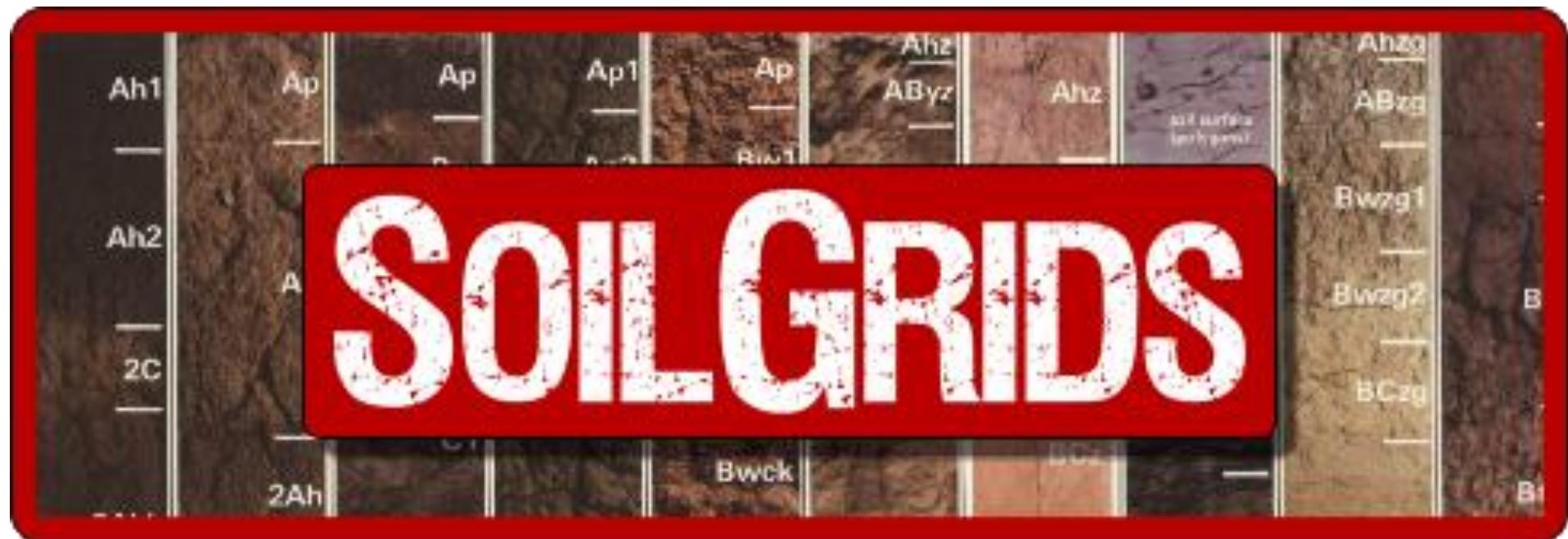
# Метеоданные

**ERA5-Land** — это климатический реанализ, глобальный набор данных о состоянии земной поверхности с высоким разрешением (~9 км). Набор данных предоставляет детальную информацию о температуре воздуха, почвы, осадках, влажности, снежном покрове и других метеорологических переменных начиная с 1950 года и до настоящего времени. Климатический реанализ — смоделированная картина состояния атмосферы, созданная путем ассимиляции всех доступных исторических наблюдений (метеостанции + ДЗЗ) с целью получения полного и согласованного набора климатических данных. ERA5-Land один из самых точных и популярных продуктов для изучения климата и гидрологических процессов на суше.



# Почвенные данные

**SoilGrids** — это глобальная цифровая почвенная карта с высоким пространственным разрешением (~250 м), которая предоставляет данные для ключевых свойств почвы (таких как содержание песка, ила, глины, органического углерода, pH) на различных глубинах. Данные SoilGrids получены на основе измерений и определений свойств в разрезах, геостатистики и машинного обучения.



# Метрики качества

**NSE (Nash-Sutcliffe Efficiency)** оценивает общую точность, сравнивая дисперсию ошибок модели с дисперсией наблюдений; её идеальное значение равно 1, а значение ниже 0 указывает, что модель хуже предсказания средним.

**KGE (Kling-Gupta Efficiency)** является сбалансированной метрикой, разлагающей точность на три компонента: корреляцию, смещение и изменчивость, и также стремясь к 1 для идеальной модели.

**RMSE (Root Mean Square Error)** показывает среднюю величину абсолютной ошибки в единицах измерения, идеал — 0, при этом метрика сильно штрафует за крупные отклонения.

**MARE (Mean Absolute Relative Error)** выражает среднюю ошибку в процентах от наблюдений, что делает её безразмерной и удобной для сравнения разных наборов данных.

**PBias (Percent Bias)** оценивает систематическую ошибку модели, показывая, насколько процентов модель в среднем завышает (положительные значения) или занижает (отрицательные значения) значения, идеал — 0%.

$$NSE = 1 - \frac{\sum_{i=1}^N (e_i - s_i)^2}{\sum_{i=1}^N (e_i - \mu(e))^2}$$

$$KGE = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (\beta-1)^2};$$

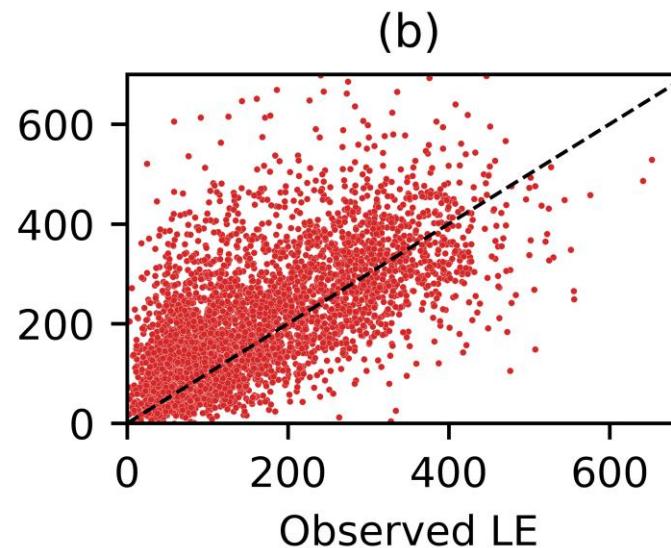
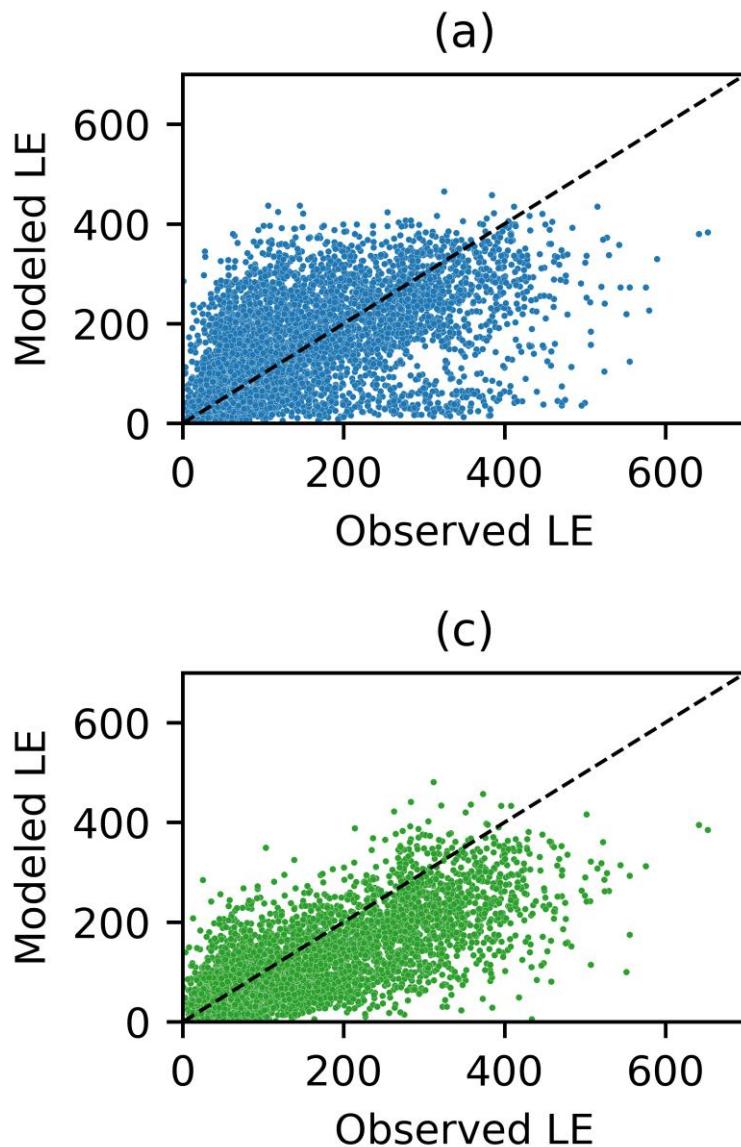
$$r = \frac{\text{cov}(e, s)}{\sigma(e) \cdot \sigma(s)}; a = \frac{\sigma(s)}{\sigma(e)}; \beta = \frac{\mu(s)}{\mu(e)}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (e_i - s_i)^2}$$

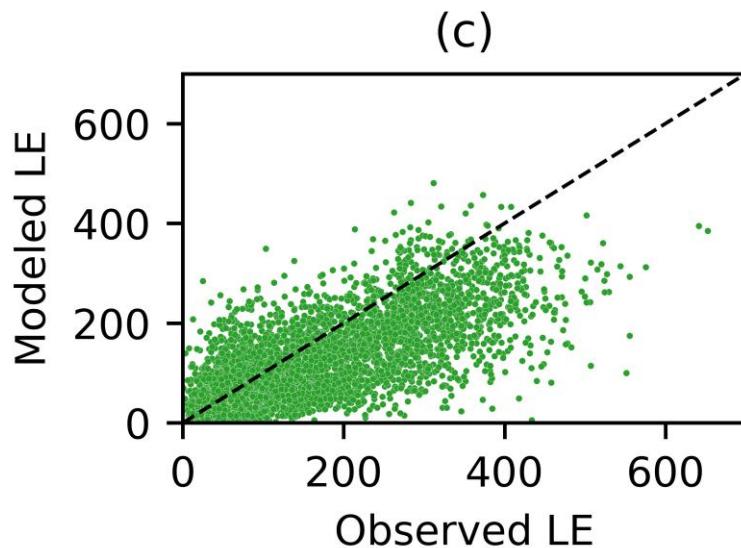
$$MARE = \frac{\sum_{i=1}^N |e_i - s_i|}{\sum_{i=1}^N e_i}$$

$$PBias = 100 \cdot \frac{\sum_{i=1}^N (e_i - s_i)}{\sum_{i=1}^N e_i}$$

# Выбор энергобалансовой модели



(a) — ERA5-Land  
(b) — TSEB  
(c) — PT-JPL

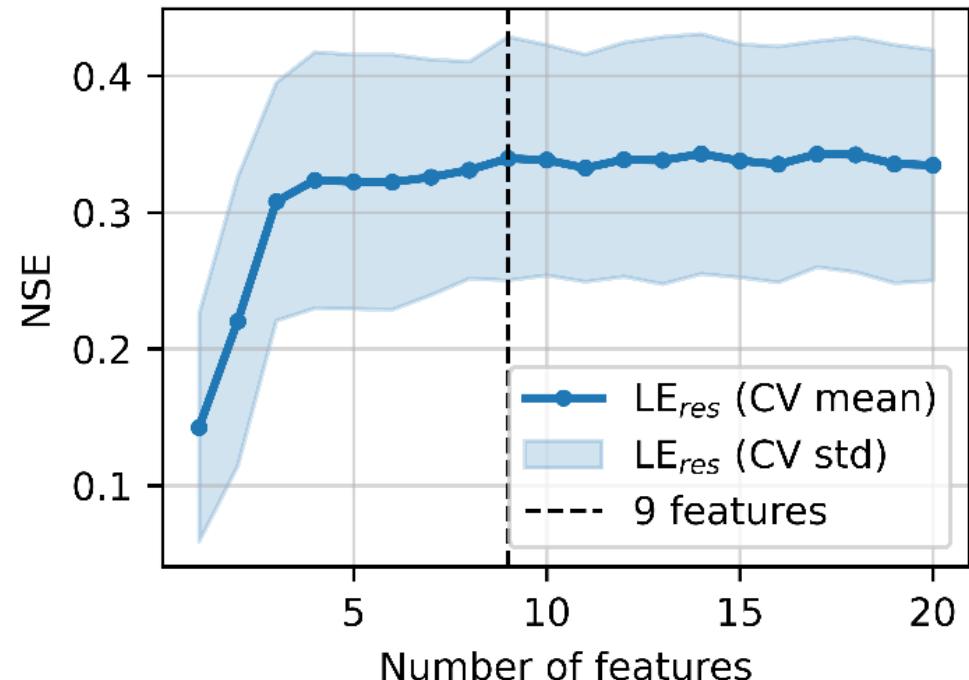


Metric	ERA5-Land	TSEB	PT-JPL
NSE	0.11	-0.05	0.39
KGE	0.50	0.52	0.55
RMSE	104.5	115.1	86.8
MARE	0.47	0.50	0.40
PBias	0.2	-26.3	24.4

# ML-ОПТИМИЗАЦИЯ МОДЕЛИ

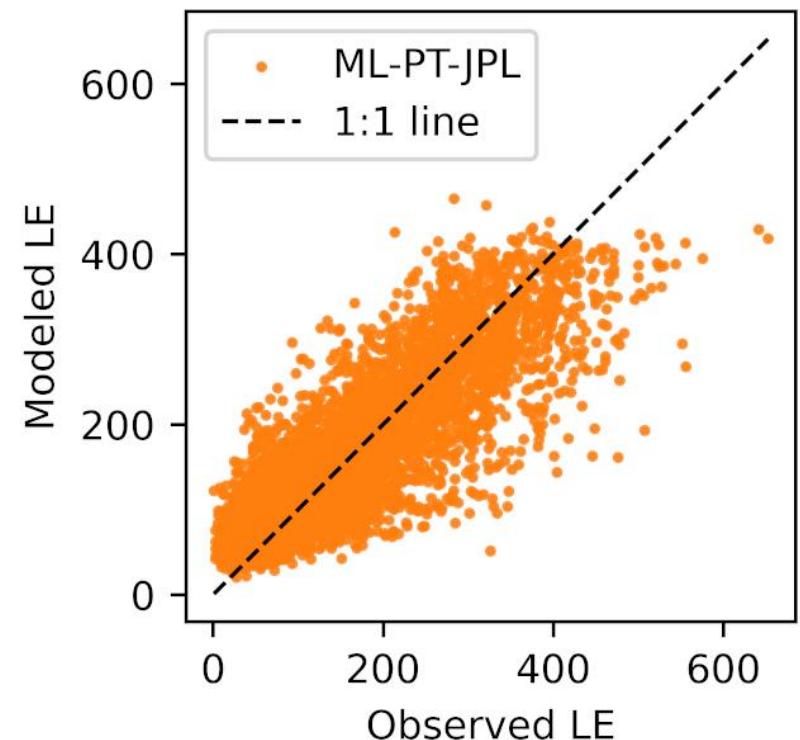
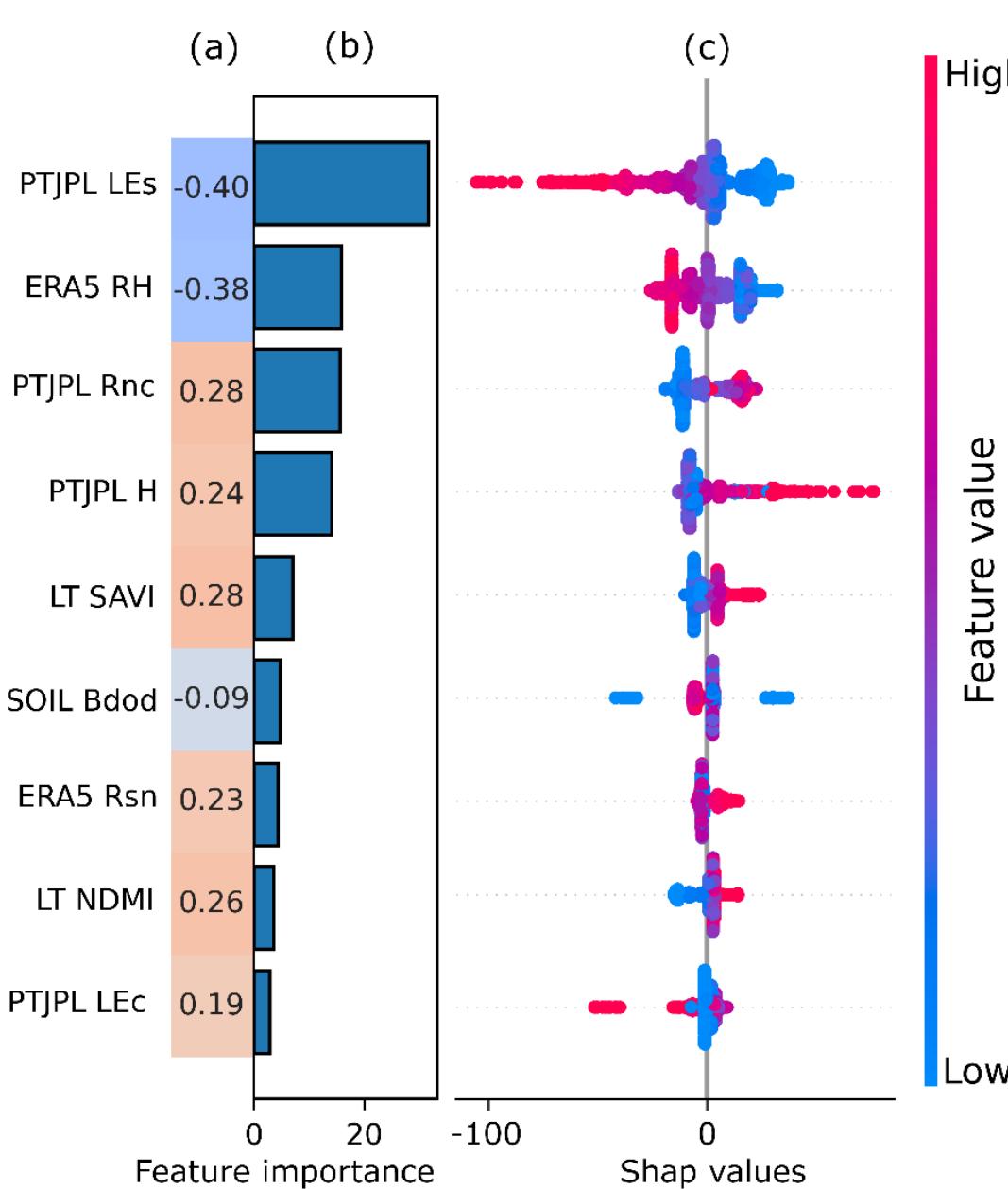
Разделение датасета на обучающую (train, 70 % случайных станций) и тестовую (test, 30 % станций) выборки.

Обучение градиентного бустинга предсказывать остатки модели PT-JPL: настройка гиперпараметров, чтобы избежать переобучения; отбор признаков на основе рекурсивного отбора с кросс-валидацией (RFECV).



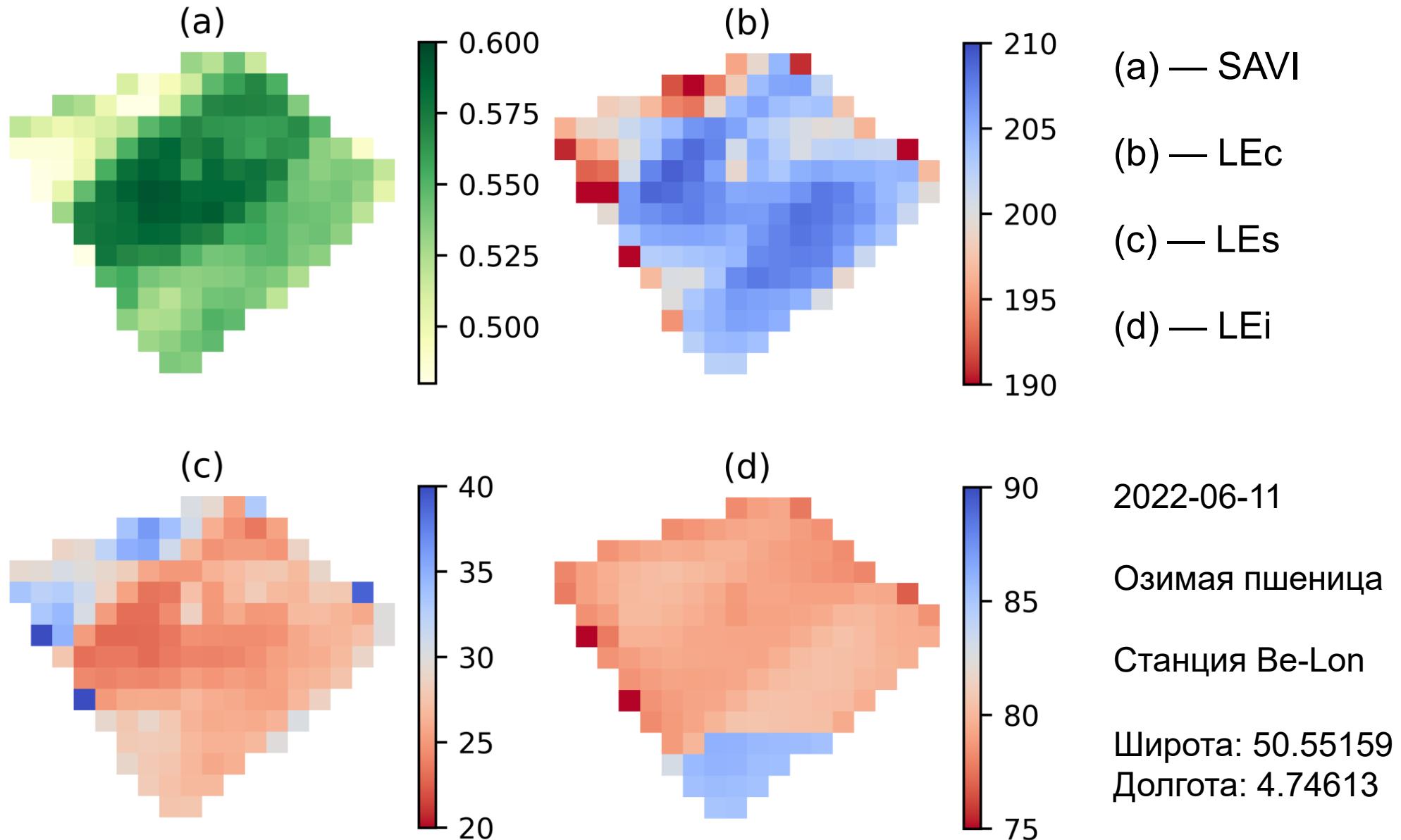
Metric	ML-PT-JPL train all features	ML-PT-JPL test all features	ML-PT-JPL train 9 features	ML-PT-JPL test 9 features
NSE	0.47	0.37	0.44	0.35
KGE	0.48	0.39	0.48	0.38
RMSE	56.0	62.3	57.7	63.3
MARE	1.04	1.29	1.07	1.32
PBias	0.0	-17.4	0.0	-20.1

# Основные факторы, влияющие на ошибку модели и метрики для ML-PT-JPL



Metric	ML-PT-JPL train	ML-PT-JPL test
NSE	0.73	0.68
KGE	0.79	0.79
RMSE	57.7	63.3
MARE	0.27	0.29
PBias	0.0	-4.4

# Распределение LE на основе ML-оптимизированной модели PT-JPL





Спасибо за  
внимание!