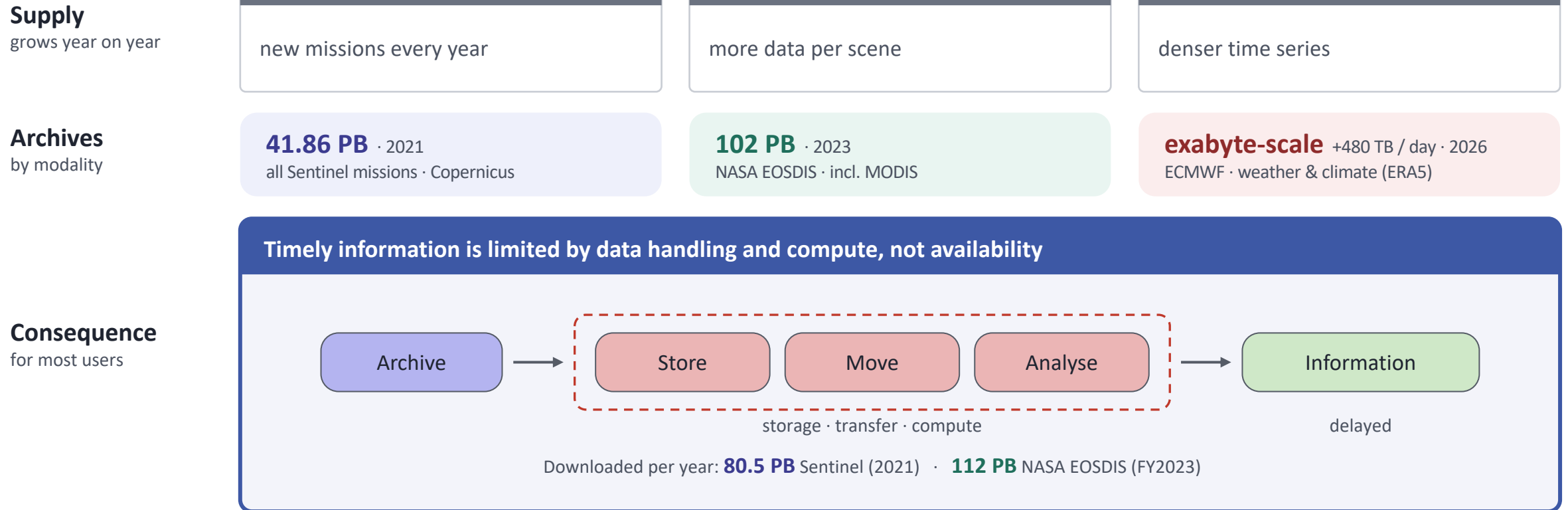


*Uncertainty-Weighted Downscaling of District Crop Yield Statistics to
Parcels Using Geospatial Foundation Model Embeddings*

The Earth-observation data problem

Satellite supply grows year on year, faster than the infrastructure to store, move and analyse it.

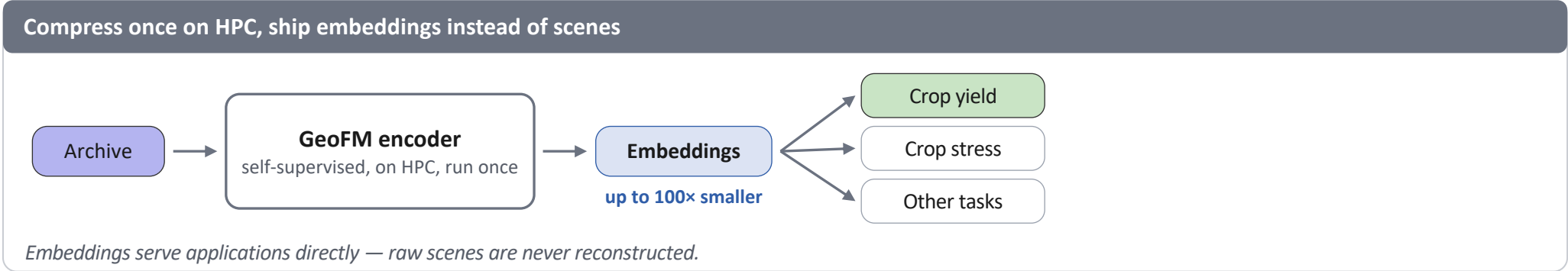


Sources: Copernicus Sentinel Data Access Annual Report 2021; NASA EOSDIS data usage metrics FY2023 (Wanchoo et al., AGU 2023); ECMWF, "The living archive" (2026).

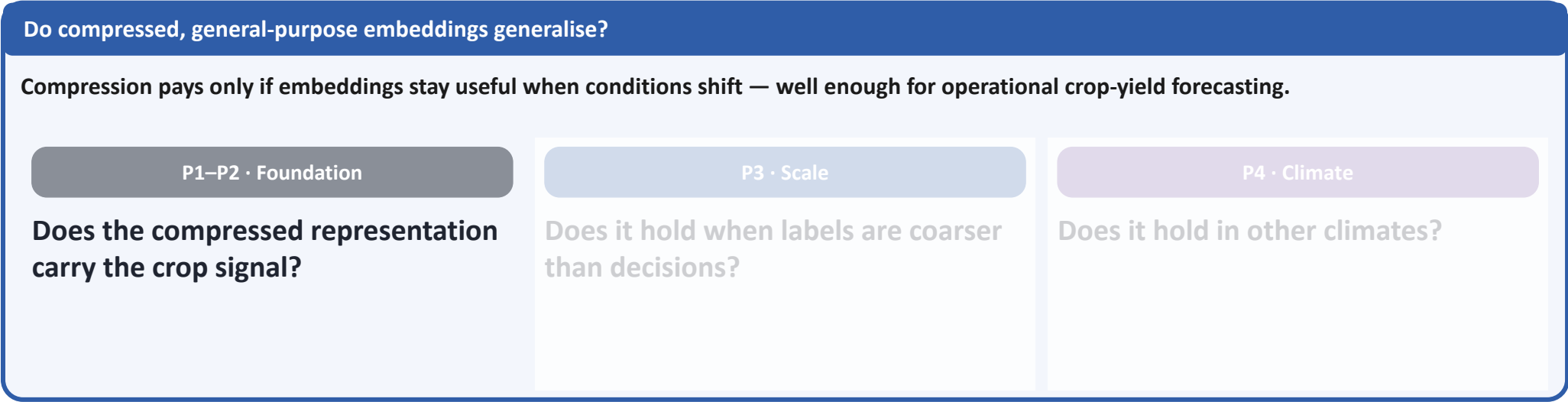
From raw data to embeddings

Embed2Scale: Compress once on HPC with self-supervised GeoFMs, then serve embeddings to applications directly.

Embed2Scale
compress once,
serve many tasks



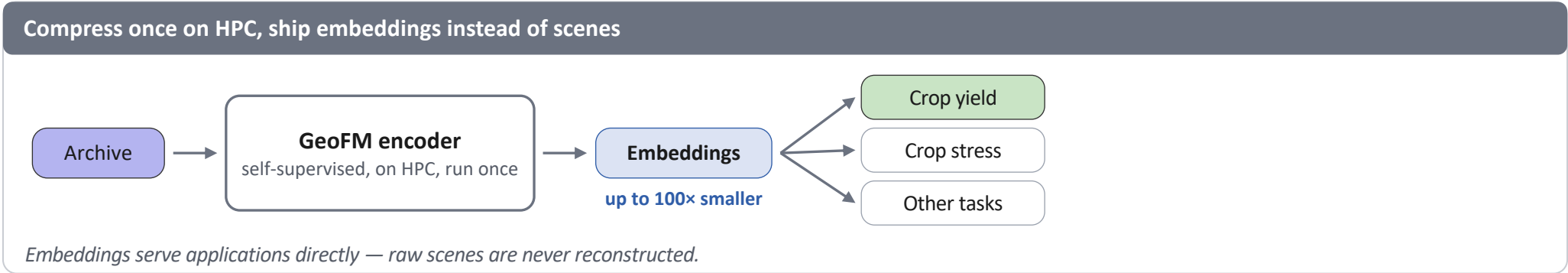
This thesis
does it transfer?



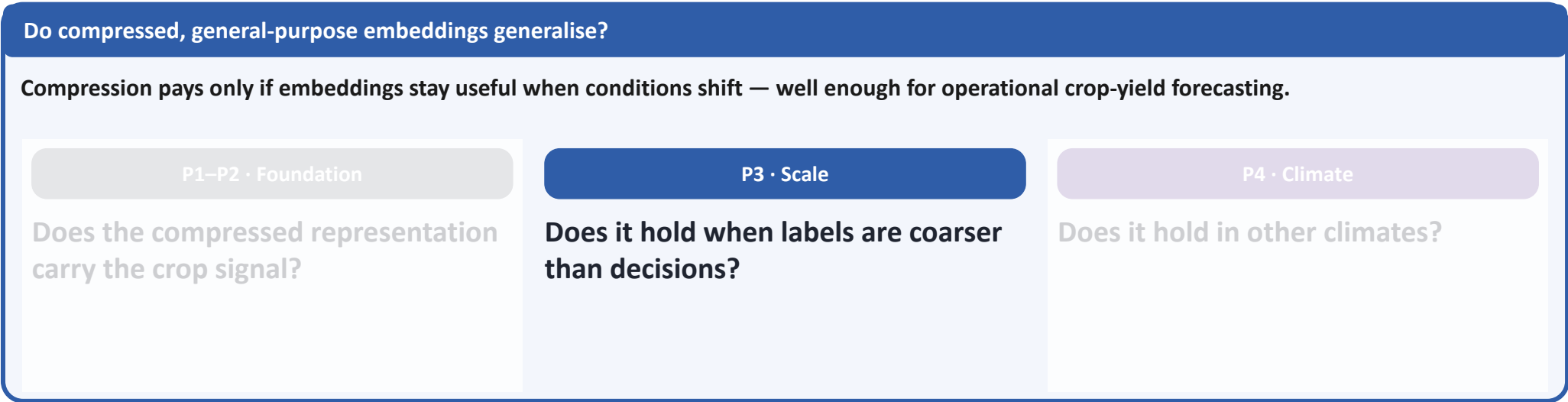
From raw data to embeddings

Embed2Scale: Compress once on HPC with self-supervised GeoFMs, then serve embeddings to applications directly.

Embed2Scale
compress once,
serve many tasks



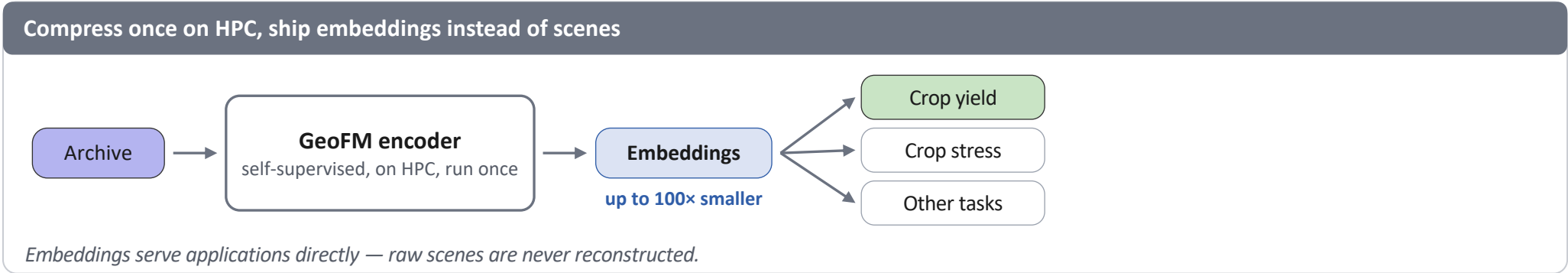
This thesis
does it transfer?



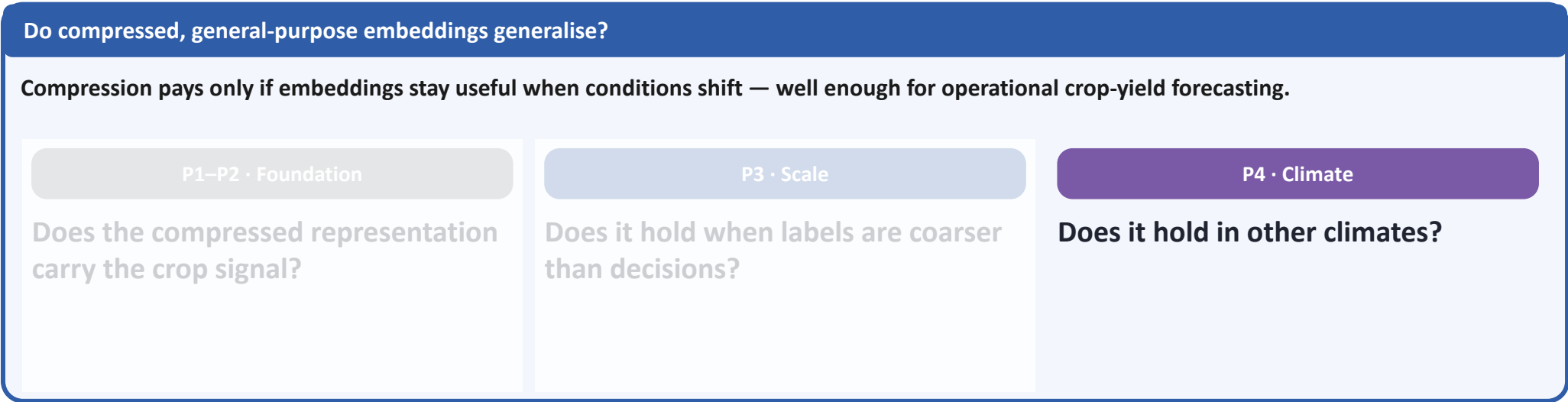
From raw data to embeddings

Embed2Scale: Compress once on HPC with self-supervised GeoFMs, then serve embeddings to applications directly.

Embed2Scale
compress once,
serve many tasks

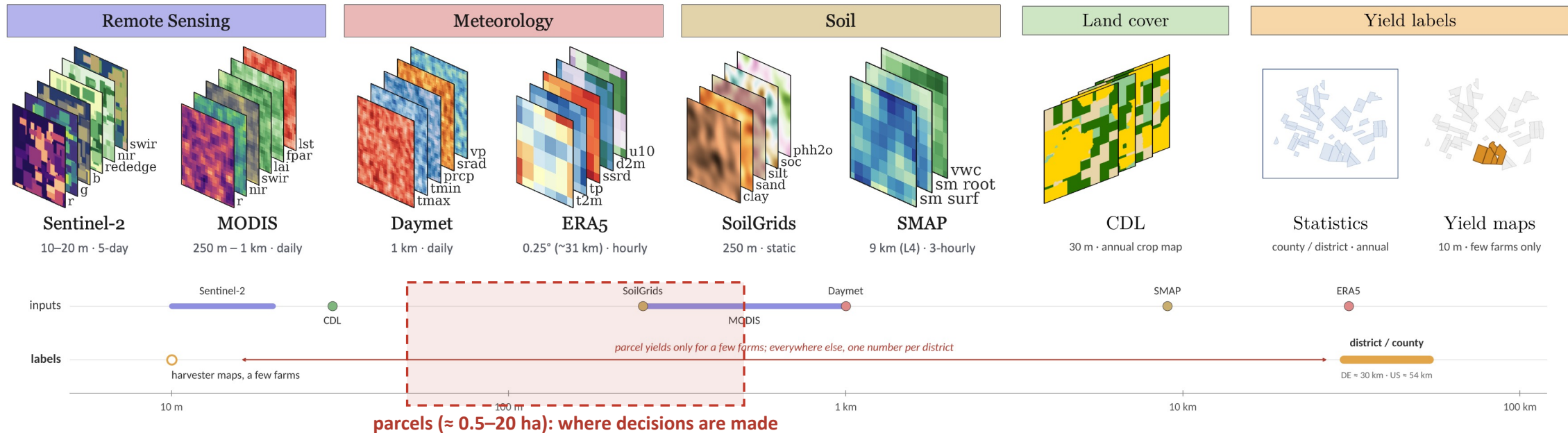


This thesis
does it transfer?



Inputs at every scale, yield labels at one

Satellite, weather and soil inputs resolve the parcel. Yield labels do not: one number per district each season, and harvester maps for a few farms.



The gap Fertiliser, insurance and stress response act on single parcels, yet yield is reported only per district. Parcel yield has to be inferred.

Thesis outline

Learning embeddings for Earth Observation applications with supercomputing

Foundation

datasets, baselines
and compressors

O1

P1

Comparison of benchmark remote sensing datasets for crop yield prediction

MODIS

Sentinel-2

Daymet

SoilGrids

CDL

USDA-NASS

O2

P2

Exploration of AI compressor methods and demo code development

Sentinel-2

AlphaEarth

Daymet

ERA5-Land

CDL

USDA-NASS

Application and scale

from validation to
deployment

O3

P3

Application of embeddings to crop yield prediction and real-world validation

Sentinel-2

EuroCrops v2

Thünen

YieldSAT

O4

P4

Scaling yield prediction for operational deployment

HLS

ERA5

NUTS-3 + county

Thünen

Datasets:

Remote sensing

Meteorology

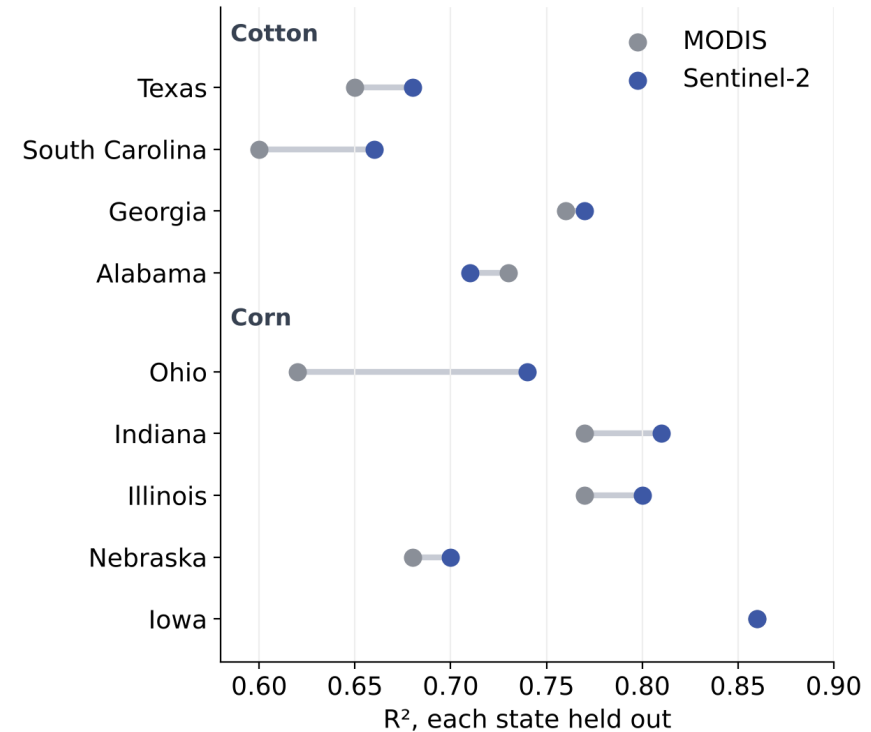
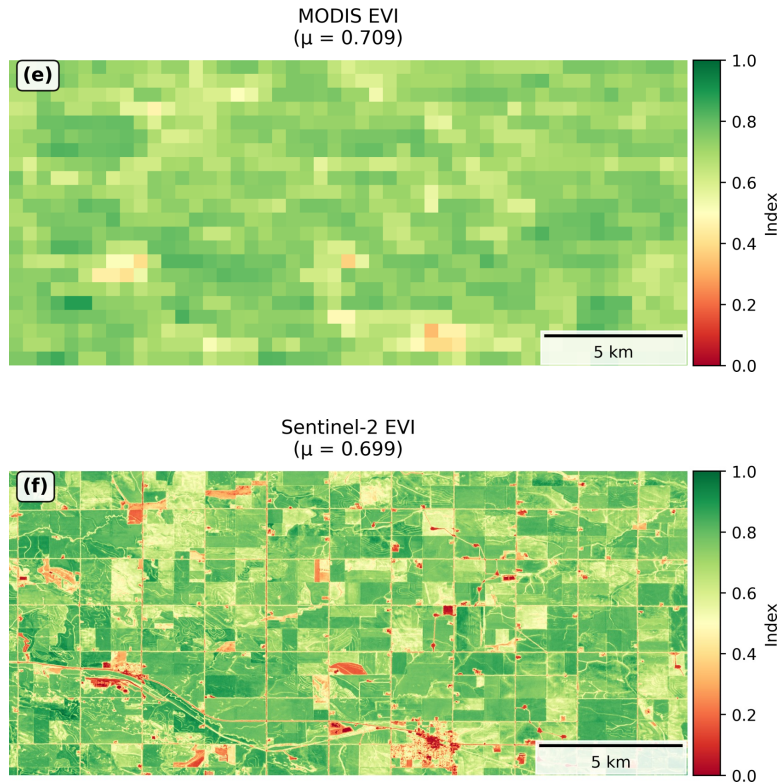
Soil

Parcels & crop masks

Yield labels

O1–O4 study objectives · P1–P4 papers

P1 · Sentinel-2's edge comes from purer crop pixels

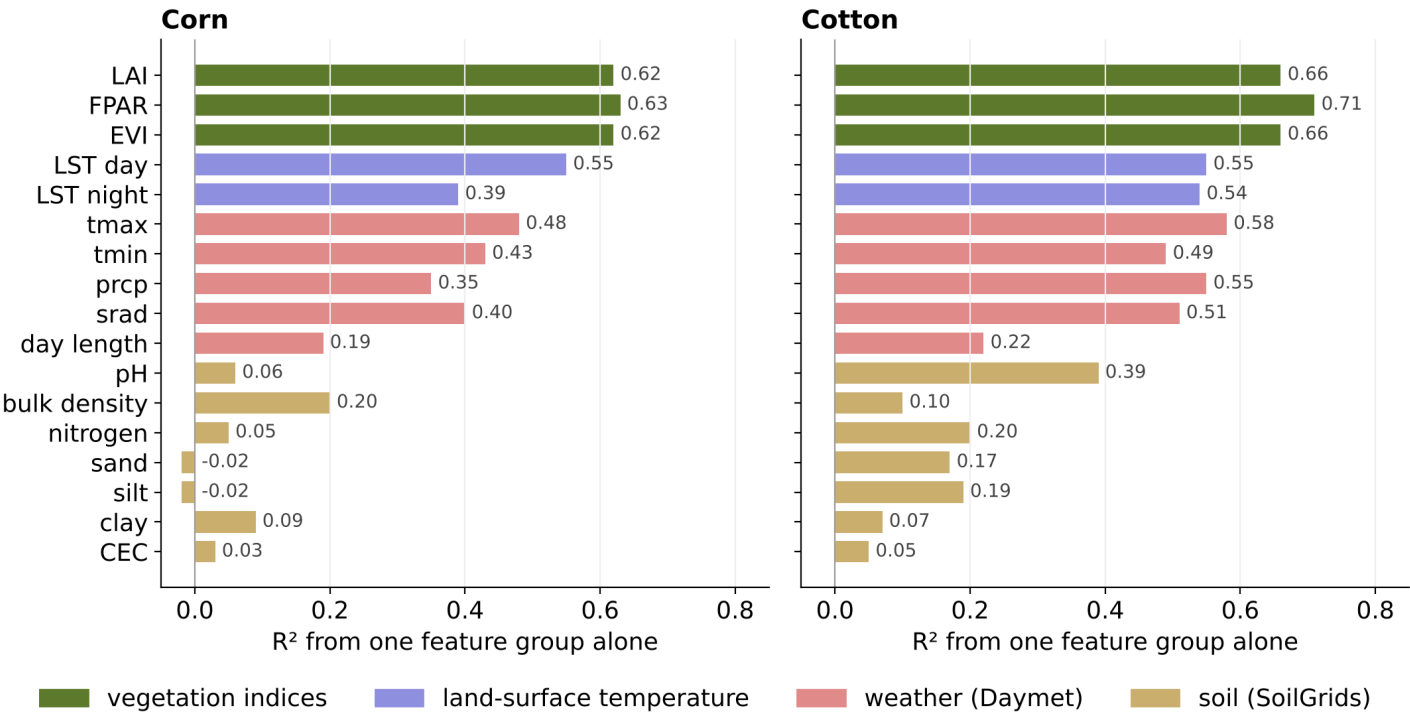


Paper I, Fig. 4

- Resolution matters more than revisit frequency for county yield. With identical inputs, models and evaluation, Sentinel-2 gave lower error than MODIS for both crops: 8.40 vs 8.69 bu/acre for corn and 123.5 vs 129.2 lb/acre for cotton (EBM).

Credits: K. Adriko, R. Sedona, L. Seguini, M. Riedel, G. Cavallaro and C. Paris, "From MODIS to Sentinel-2: A Regional Comparative Analysis of Crop-Yield Prediction With Matched Spatiotemporal Data," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 27663-27683, 2025, doi: 10.1109/JSTARS.2025.3624046.

P1 · The yield signal is predominantly in the vegetation dynamics



- The yield signal sits in vegetation dynamics. Explainable AI (SHAP on EBM) ranked EVI, FPAR and LAI as the leading predictors for both sensors.

Credits: K. Adriko, R. Sedona, L. Seguini, M. Riedel, G. Cavallaro and C. Paris, "From MODIS to Sentinel-2: A Regional Comparative Analysis of Crop-Yield Prediction With Matched Spatiotemporal Data," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 27663-27683, 2025, doi: 10.1109/JSTARS.2025.3624046.

Thesis Outline

Frozen GeoFM embeddings for crop yield: four study objectives, delivered through a series of papers. Defence targeted Q4 2027.

Foundation

datasets, baselines
and compressors

O1

P1

Comparison of benchmark remote sensing datasets for crop yield prediction

MODIS

Sentinel-2

Daymet

SoilGrids

CDL

USDA-NASS

O2

P2

Exploration of AI compressor methods and demo code development

Sentinel-2

AlphaEarth

Daymet

ERA5-Land

CDL

USDA-NASS

Application and scale

from validation to
deployment

O3

P3

Application of embeddings to crop yield prediction and real-world validation

Sentinel-2

EuroCrops v2

Thünen

YieldSAT

O4

P4

Scaling yield prediction for operational deployment

HLS

ERA5

NUTS-3 + county

Thünen

Datasets:

Remote sensing

Meteorology

Soil

Parcels & crop masks

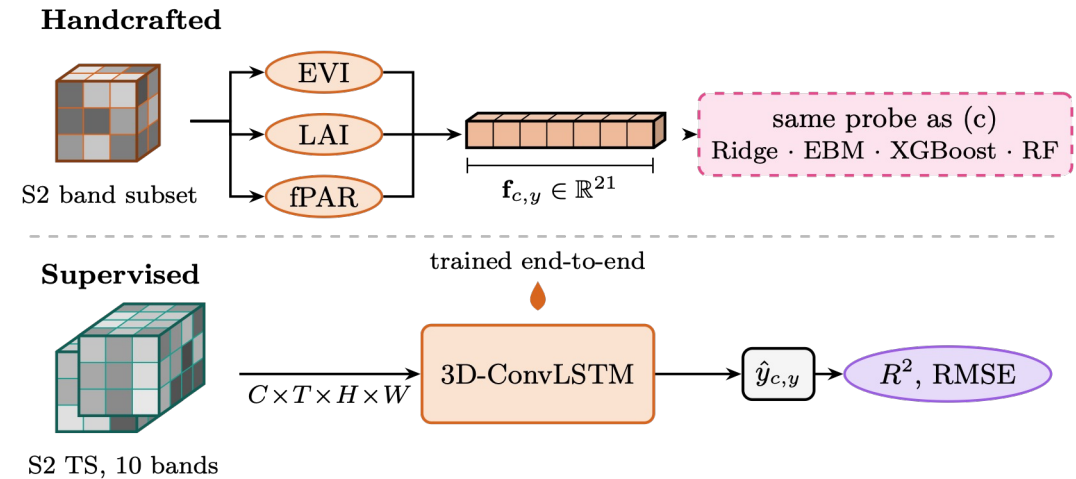
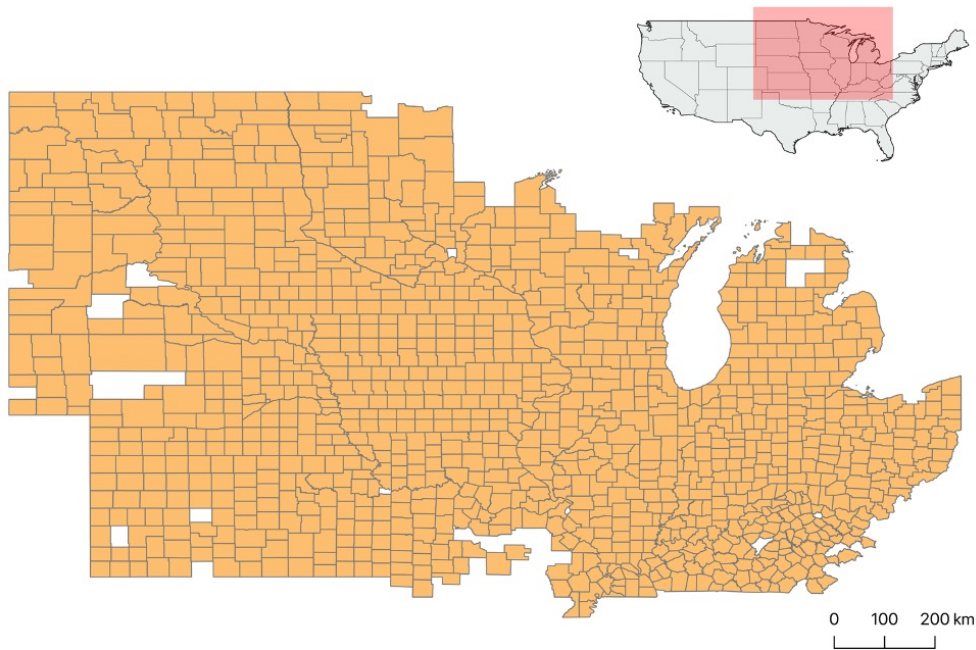
Yield labels

O1–O4 study objectives · P1–P4 papers

P2. Exploration of AI compressor methods and demo code development

REGIONAL US Corn Belt

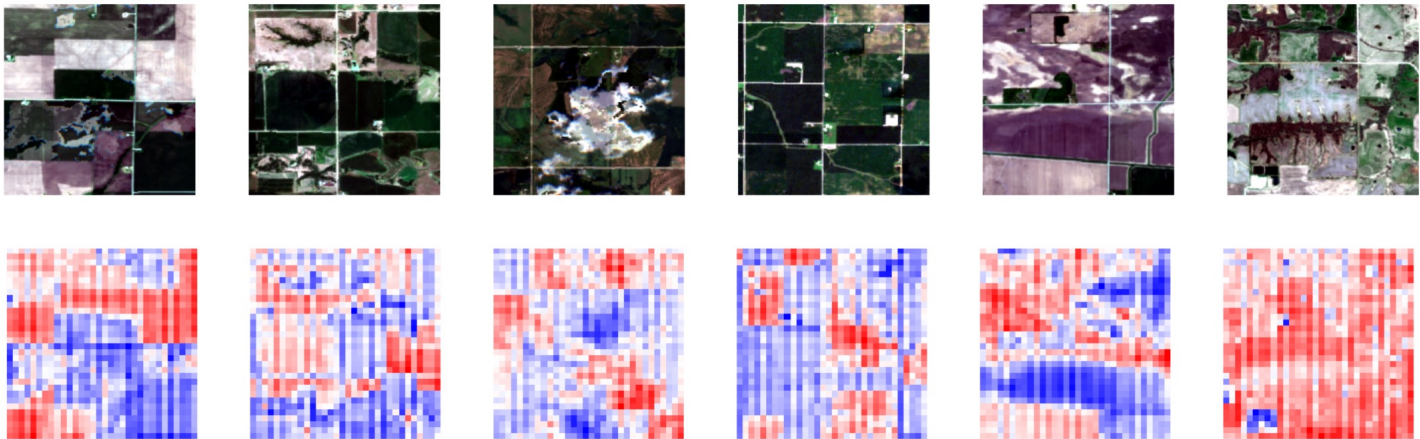
County-level



Most GeoFM evaluations target spatial tasks (segmentation, classification). Temporal tasks such as yield estimation remain largely untested.

Credits: K. Adriko, R. Sedona, M. Riedel, G. Cavallaro and C. Paris, "From Pretraining to Prediction: Benchmarking Geospatial Foundation Models for Corn Yield Estimation," submitted to IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, in revision, 2026.

P2. Exploration of AI compressor methods and demo code development



Top: Sentinel-2 patches (RGB)
Bottom: frozen Clay embeddings of the same patches, shown as 2-D maps
Field boundaries and land-cover transitions remain visible in the embedding

Paper II, Fig. 4

Pretrained with self-supervision; embeddings computed once on HPC, stored, reused

273 GB → 2.63 GB

Compressed Sentinel-2 patches vs stored Clay embeddings (Paper II)

0.47–0.93 GPU h

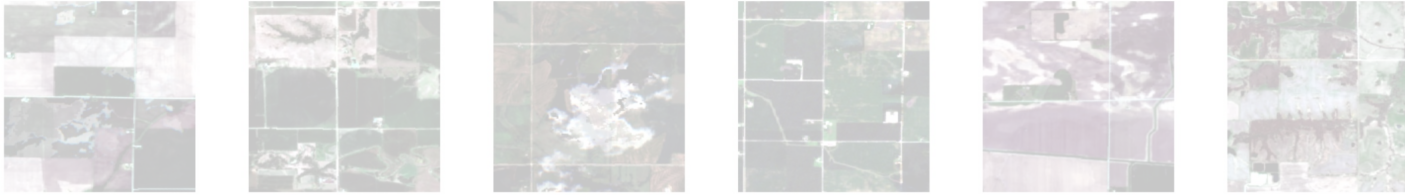
Embedding extraction per encoder, full corpus, one GH200 node

Open question

Do embeddings keep yield-relevant phenology, and does it transfer?

Credits: K. Adriko, R. Sedona, M. Riedel, G. Cavallaro and C. Paris, “From Pretraining to Prediction: Benchmarking Geospatial Foundation Models for Corn Yield Estimation,” submitted to IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, in revision, 2026.

P2. Exploration of AI compressor methods and demo code development



Top: Sentinel-2 patches (RGB)

Bottom: frozen Clay embeddings of the same patches, shown as 2-D maps

Do embeddings keep yield-relevant phenology, and does it transfer?

273 GB → 2.63 GB

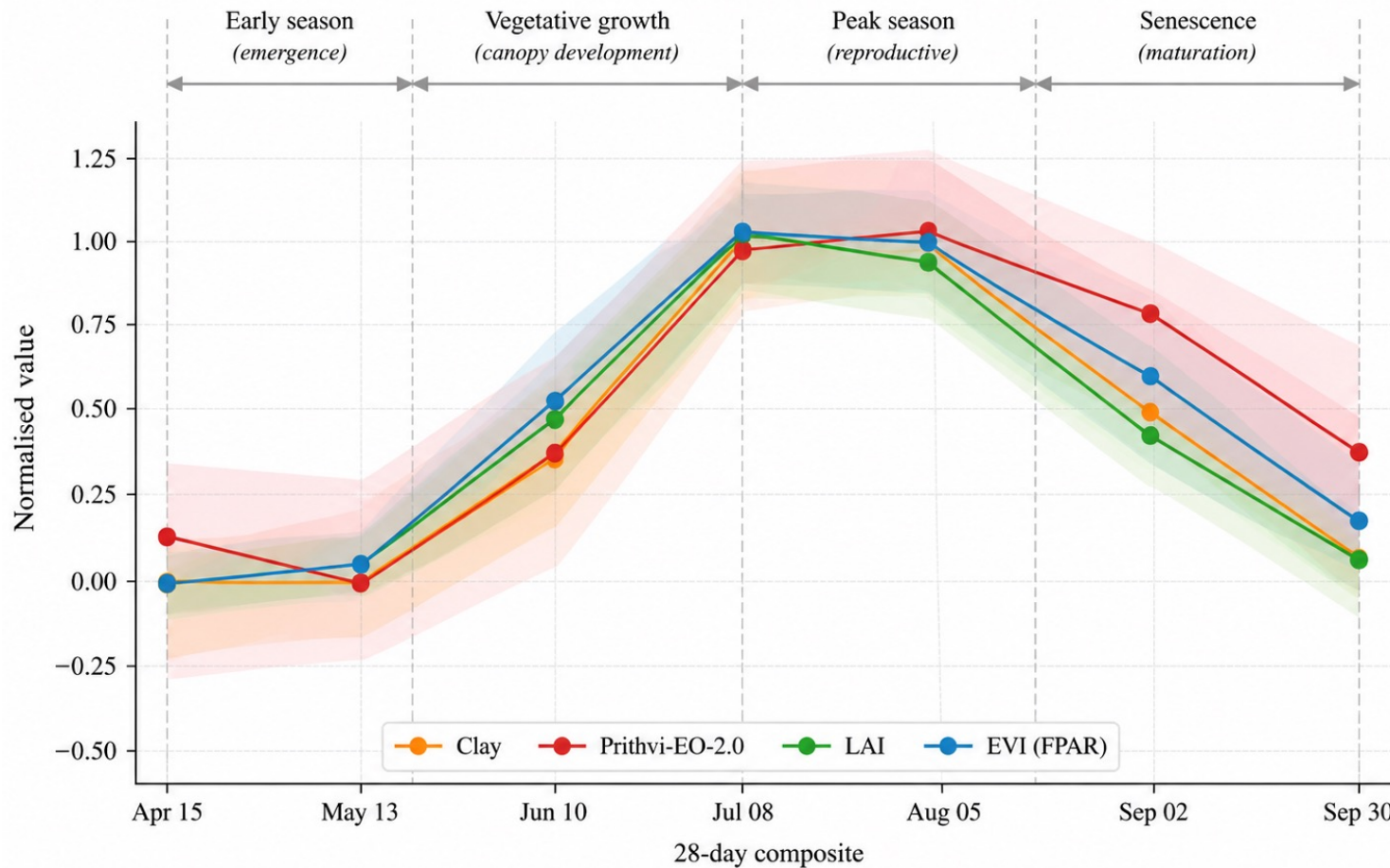
0.47–0.93 GPU h

Compressed Sentinel-2 patches vs stored Clay embeddings (Paper II)

Embedding extraction per encoder, full corpus, one GH200 node

Credits: K. Adriko, R. Sedona, M. Riedel, G. Cavallaro and C. Paris, “From Pretraining to Prediction: Benchmarking Geospatial Foundation Models for Corn Yield Estimation,” submitted to IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, in revision, 2026.

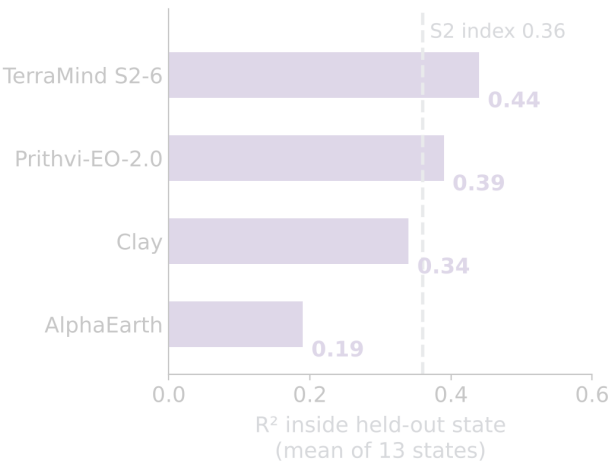
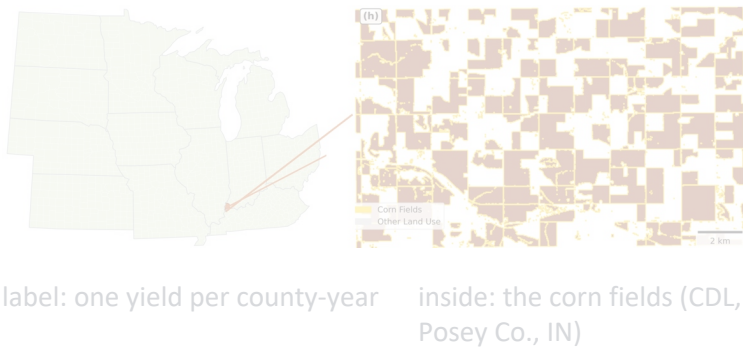
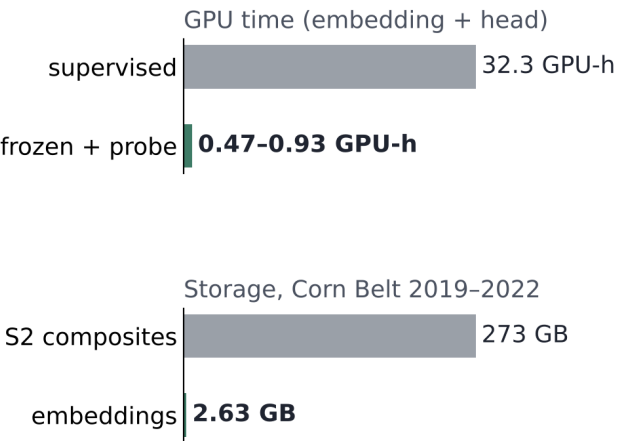
P2 · Frozen embeddings track crop phenology



- PC1 of Clay and Prithvi-EO-2.0 embeddings vs vegetation indices over seven 28-day composites, mid-April to end-September
- Embeddings follow crop phenology: rising through canopy development (timesteps 2–4), peaking in July–August, declining in senescence
- These curves become the no-stress reference for stress detection

Credits: K. Adriko, R. Sedona, M. Riedel, G. Cavallaro and C. Paris, "From Pretraining to Prediction: Benchmarking Geospatial Foundation Models for Corn Yield Estimation," submitted to IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, in revision, 2026.

P2 · Takeaway and two limitations



Frozen embeddings reach the accuracy of a supervised 3D-ConvLSTM (linear probe R^2 0.78–0.82 vs 0.80; S2 index 0.77) with 35–70× less GPU time and ~100× less storage.

Basis for frozen TerraMind in P3 and P4.

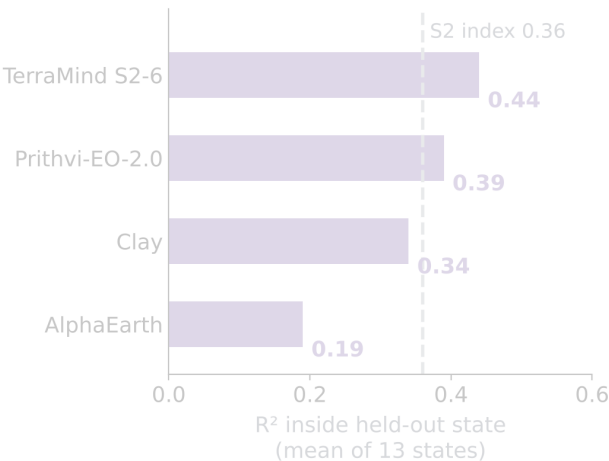
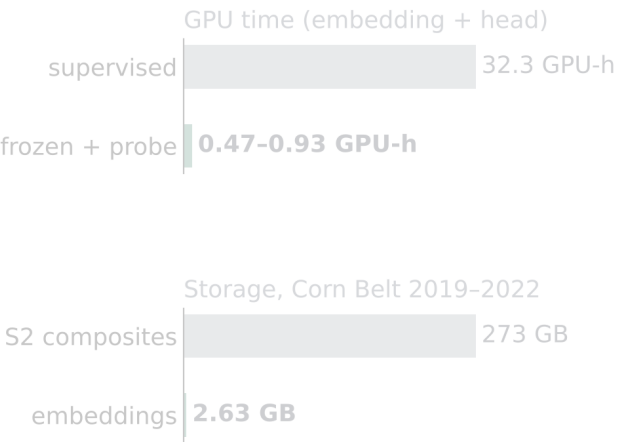
One label per county-year. P2 cannot say anything about the fields inside a county, where agricultural decisions are made.

P3: parcels, with the district statistic fixing the level.

With a whole state held out, R^2 inside that state falls to 0.19–0.44. No embedding is clearly above the Sentinel-2 index.

P4: climate-conditioned transfer across regions.

P2 · Takeaway and two limitations

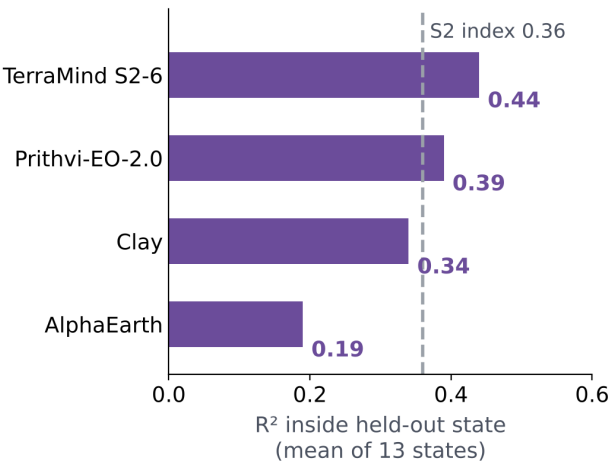
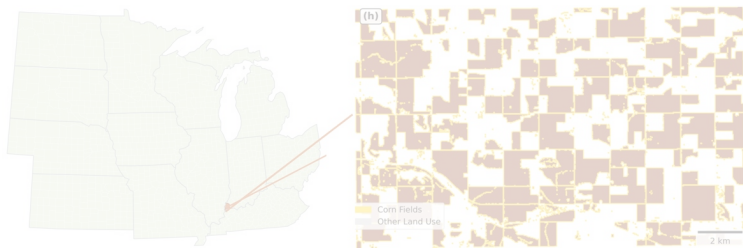
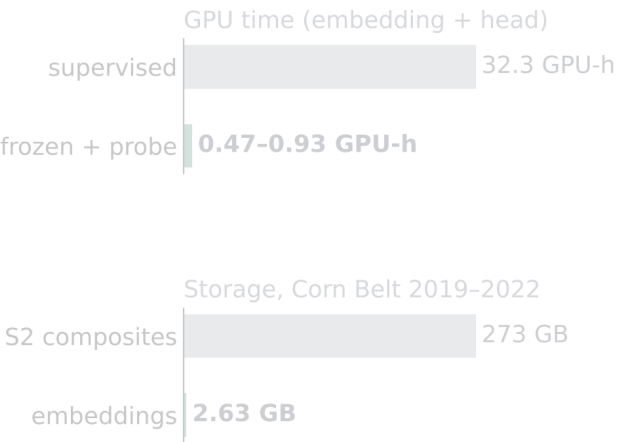


Frozen embeddings reach the accuracy of a supervised 3D-ConvLSTM (linear probe R^2 0.78–0.82 vs 0.80; S2 index 0.77) with 35–70× less GPU time and ~100× less storage.
Basis for frozen TerraMind in P3 and P4.

One label per county-year. P2 cannot say anything about the fields inside a county, where agricultural decisions are made.
P3: parcels, with the district statistic fixing the level.

With a whole state held out, R^2 inside that state falls to 0.19–0.44. No embedding is clearly above the Sentinel-2 index.
P4: climate-conditioned transfer across regions.

P2 · Takeaway and two limitations



Frozen embeddings reach the accuracy of a supervised 3D-ConvLSTM (linear probe R^2 0.78–0.82 vs 0.80; S2 index 0.77) with 35–70× less GPU time and ~100× less storage.
Basis for frozen TerraMind in P3 and P4.

One label per county-year. P2 cannot say anything about the fields inside a county, where agricultural decisions are made.
P3: parcels, with the district statistic fixing the level.

With a whole state held out, R^2 inside that state falls to 0.19–0.44. No embedding is clearly above the Sentinel-2 index.
P4: climate-conditioned transfer across regions.

Credits: K. Adriko, R. Sedona, M. Riedel, G. Cavallaro and C. Paris, “From Pretraining to Prediction: Benchmarking Geospatial Foundation Models for Corn Yield Estimation,” submitted to IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, in revision, 2026.

Thesis Outline

Learning embeddings for Earth Observation applications with supercomputing.

Foundation

datasets, baselines
and compressors

O1

P1

Comparison of benchmark remote sensing datasets for crop yield prediction

MODIS

Sentinel-2

Daymet

SoilGrids

CDL

USDA-NASS

O2

P2

Exploration of AI compressor methods and demo code development

Sentinel-2

AlphaEarth

Daymet

ERA5-Land

CDL

USDA-NASS

Application and scale

from validation to
deployment

O3

P3

Application of embeddings to crop yield prediction and real-world validation

Sentinel-2

EuroCrops v2

Thünen

YieldSAT

O4

P4

Scaling yield prediction for operational deployment

HLS

ERA5

NUTS-3 + county

Thünen

Datasets:

Remote sensing

Meteorology

Soil

Parcels & crop masks

Yield labels

O1–O4 study objectives · P1–P4 papers · colours follow the Embed2Scale project deck

P3 · Why embeddings, and why scale

WHY EMBEDDINGS

Every parcel, every season.

Computed once on HPC from Sentinel-2, so the input exists wherever wheat grows, not only where yield was measured.

Few labels needed.

The encoder stays frozen; only a small head is trained. That matters when parcel labels are 132 field-seasons from six farms.

Shown to carry crop signal.

P2: they track phenology without labels and match a supervised model at 35–70× less GPU time.

WHY SCALE

Decisions are per parcel.

The district statistic averages away the variation that management, insurance and stress response act on.

Labels stop at the district.

Statistics are complete and unbiased but coarse; field yields are precise but rare. Neither alone gives a parcel map.

The thesis test (O3).

Compression pays only if embeddings stay useful when the scale changes: county in P2, parcel in P3.

P3 Embeddings give every parcel an input; the district statistic gives every district a label. Joining them yields parcel estimates with uncertainty.

P3 · Why embeddings, and why scale

WHY EMBEDDINGS

Every parcel, every season.

Computed once on HPC from Sentinel-2, so the input exists wherever wheat grows, not only where yield was measured.

Few labels needed.

The encoder stays frozen; only a small head is trained. That matters when parcel labels are 132 field-seasons from six farms.

Shown to carry crop signal.

P2: they track phenology without labels and match a supervised model at 35–70× less GPU time.

WHY SCALE

Decisions are per parcel.

The district statistic averages away the variation that management, insurance and stress response act on.

Labels stop at the district.

Statistics are complete and unbiased but coarse; field yields are precise but rare. Neither alone gives a parcel map.

The thesis test (O3).

Compression pays only if embeddings stay useful when the scale changes: county in P2, parcel in P3.

P3 Embeddings give every parcel an input; the district statistic gives every district a label. Joining them yields parcel estimates with uncertainty.

P3 · Why embeddings, and why scale

WHY EMBEDDINGS

Every parcel, every season.

Computed once on HPC from Sentinel-2, so the input exists wherever wheat grows, not only where yield was measured.

Few labels needed.

The encoder stays frozen; only a small head is trained. That matters when parcel labels are 132 field-seasons from six farms.

Shown to carry crop signal.

P2: they track phenology without labels and match a supervised model at 35–70× less GPU time.

WHY SCALE

Decisions are per parcel.

The district statistic averages away the variation that management, insurance and stress response act on.

Labels stop at the district.

Statistics are complete and unbiased but coarse; field yields are precise but rare. Neither alone gives a parcel map.

The thesis test (O3).

Compression pays only if embeddings stay useful when the scale changes: county in P2, parcel in P3.

P3 Embeddings give every parcel an input; the district statistic gives every district a label. Joining them yields parcel estimates with uncertainty.

P3 · Why embeddings, and why scale

WHY EMBEDDINGS

Every parcel, every season.

Computed once on HPC from Sentinel-2, so the input exists wherever wheat grows, not only where yield was measured.

Few labels needed.

The encoder stays frozen; only a small head is trained. That matters when parcel labels are 132 field-seasons from six farms.

Shown to carry crop signal.

P2: they track phenology without labels and match a supervised model at 35–70× less GPU time.

WHY SCALE

Decisions are per parcel.

The district statistic averages away the variation that management, insurance and stress response act on.

Labels stop at the district.

Statistics are complete and unbiased but coarse; field yields are precise but rare. Neither alone gives a parcel map.

The thesis test (O3).

Compression pays only if embeddings stay useful when the scale changes: county in P2, parcel in P3.

P3 Embeddings give every parcel an input; the district statistic gives every district a label. Joining them yields parcel estimates with uncertainty.

P3 · Why embeddings, and why scale

WHY EMBEDDINGS

Every parcel, every season.

Computed once on HPC from Sentinel-2, so the input exists wherever wheat grows, not only where yield was measured.

Few labels needed.

The encoder stays frozen; only a small head is trained. That matters when parcel labels are 132 field-seasons from six farms.

Shown to carry crop signal.

P2: they track phenology without labels and match a supervised model at 35–70× less GPU time.

WHY SCALE

Decisions are per parcel.

The district statistic averages away the variation that management, insurance and stress response act on.

Labels stop at the district.

Statistics are complete and unbiased but coarse; field yields are precise but rare. Neither alone gives a parcel map.

The thesis test (O3).

Compression pays only if embeddings stay useful when the scale changes: county in P2, parcel in P3.

P3 Embeddings give every parcel an input; the district statistic gives every district a label. Joining them yields parcel estimates with uncertainty.

P3 · Why embeddings, and why scale

WHY EMBEDDINGS

Every parcel, every season.

Computed once on HPC from Sentinel-2, so the input exists wherever wheat grows, not only where yield was measured.

Few labels needed.

The encoder stays frozen; only a small head is trained. That matters when parcel labels are 132 field-seasons from six farms.

Shown to carry crop signal.

P2: they track phenology without labels and match a supervised model at 35–70× less GPU time.

WHY SCALE

Decisions are per parcel.

The district statistic averages away the variation that management, insurance and stress response act on.

Labels stop at the district.

Statistics are complete and unbiased but coarse; field yields are precise but rare. Neither alone gives a parcel map.

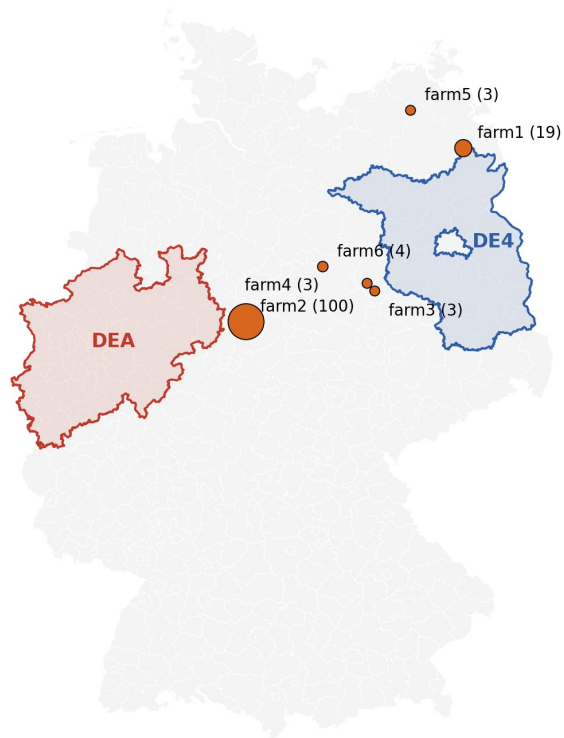
The thesis test (O3).

Compression pays only if embeddings stay useful when the scale changes: county in P2, parcel in P3.

P3 Embeddings give every parcel an input; the district statistic gives every district a label. Joining them yields parcel estimates with uncertainty.

P3. Application of embeddings to crop yield prediction and real-world validation

YieldSAT Germany: 132 winter-wheat field-seasons, 2018–2024
marker = farm, label = field-seasons



Sentinel-2 L2A

10 bands; 5 windows Oct–mid-Aug anchored on a green-up date G from imagery; GEE harmonised S2 SR

TerraMind v1-base

frozen; 16 × 16-px tokens pooled into each parcel by purity weight → 5 × 768 + validity mask

EuroCrops v2

winter wheat (code 115); 2020: 8,661 parcels in DE4, 65,153 in DEA. Parcel → district by point-on-surface; eligible ≥ 0.5 ha, area within 10 % of declared

Thünen districts

winter-wheat yield, t/ha, 2019–2021; a district-season is kept if mapped / reported wheat area ≥ 0.7 — DE4 14 of 18, DEA 29 of 53

YieldSAT fields

132 wheat field-seasons, 6 farms, 2018–2024; field = valid yield-map pixels; all outside DE4 and DEA

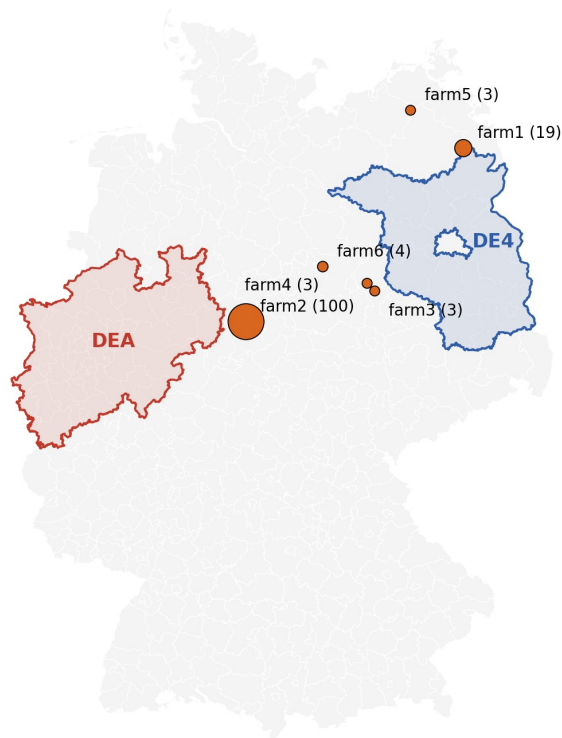
SPLIT

Train: DE4 2019–2020 statistics + fields 2018–2020. Test: DE4 2021 (time), DEA 2019–2020 (space), DEA 2021 (both), fields 2021.

Credits: Miranda, M., Pathak, D., Helber, P., Bischke, B., Najjar, H., Mena, F., ... & Dengel, A. (2026). Yieldsat: A multimodal benchmark dataset for high-resolution crop yield prediction. arXiv preprint arXiv:2604.00940.

P3. Application of embeddings to crop yield prediction and real-world validation

YieldSAT Germany: 132 winter-wheat field-seasons, 2018–2024
marker = farm, label = field-seasons



Sentinel-2 L2A

10 bands; 5 windows Oct–mid-Aug anchored on a green-up date G from imagery; GEE harmonised S2 SR

TerraMind v1-base

frozen; 16 × 16-px tokens pooled into each parcel by purity weight → 5 × 768 + validity mask

EuroCrops v2

winter wheat (code 115); 2020: 8,661 parcels in DE4, 65,153 in DEA. Parcel → district by point-on-surface; eligible ≥ 0.5 ha, area within 10 % of declared

Thünen districts

winter-wheat yield, t/ha, 2019–2021; a district-season is kept if mapped / reported wheat area ≥ 0.7 — DE4 14 of 18, DEA 29 of 53

YieldSAT fields

132 wheat field-seasons, 6 farms, 2018–2024; field = valid yield-map pixels; all outside DE4 and DEA

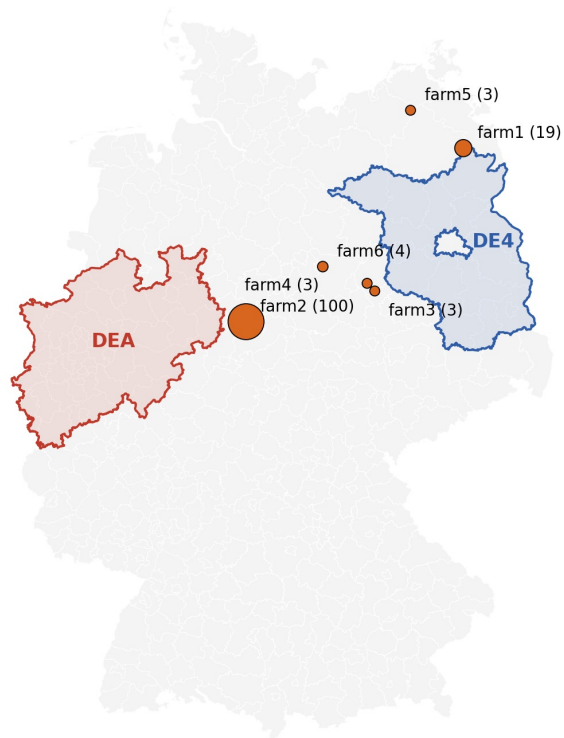
SPLIT

Train: DE4 2019–2020 statistics + fields 2018–2020. Test: DE4 2021 (time), DEA 2019–2020 (space), DEA 2021 (both), fields 2021.

Credits: Miranda, M., Pathak, D., Helber, P., Bischke, B., Najjar, H., Mena, F., ... & Dengel, A. (2026). Yieldsat: A multimodal benchmark dataset for high-resolution crop yield prediction. arXiv preprint arXiv:2604.00940.

P3. Application of embeddings to crop yield prediction and real-world validation

YieldSAT Germany: 132 winter-wheat field-seasons, 2018–2024
marker = farm, label = field-seasons



Sentinel-2 L2A

10 bands; 5 windows Oct–mid-Aug anchored on a green-up date G from imagery; GEE harmonised S2 SR

TerraMind v1-base

frozen; 16 × 16-px tokens pooled into each parcel by purity weight → 5 × 768 + validity mask

EuroCrops v2

winter wheat (code 115); 2020: 8,661 parcels in DE4, 65,153 in DEA. Parcel → district by point-on-surface; eligible ≥ 0.5 ha, area within 10 % of declared

Thünen districts

winter-wheat yield, t/ha, 2019–2021; a district-season is kept if mapped / reported wheat area ≥ 0.7 — DE4 14 of 18, DEA 29 of 53

YieldSAT fields

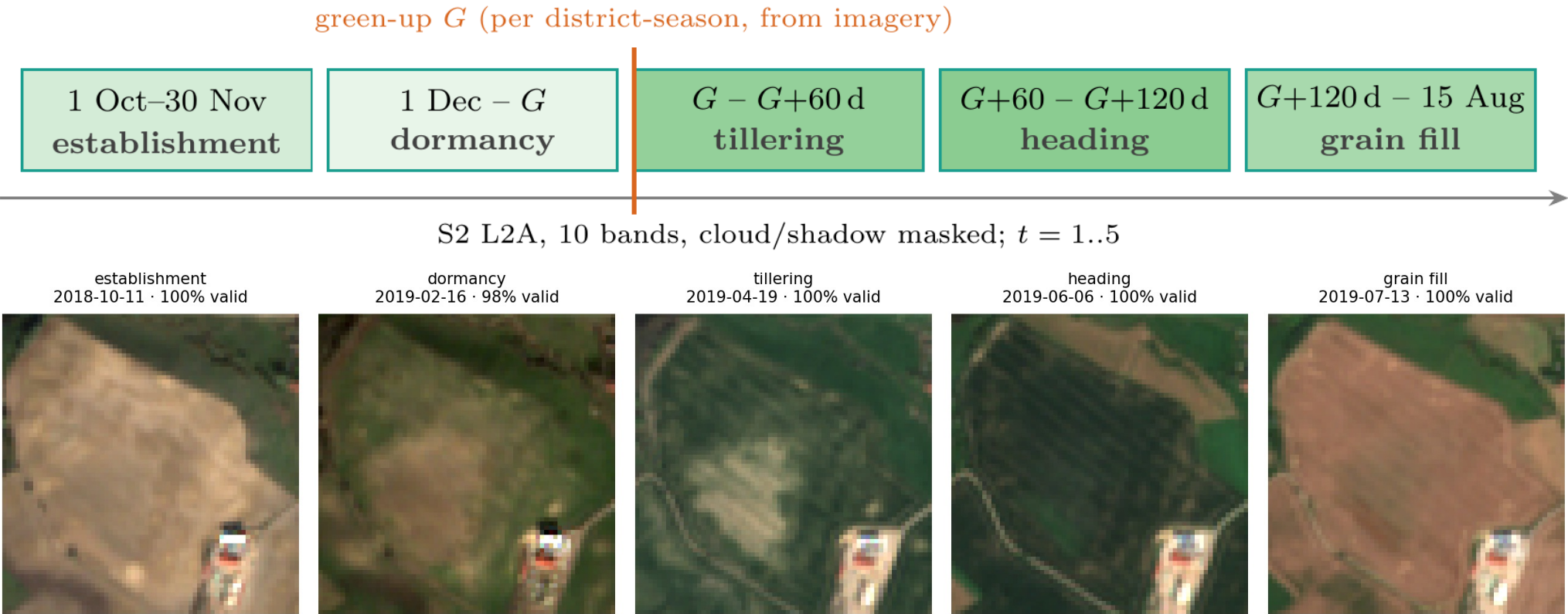
132 wheat field-seasons, 6 farms, 2018–2024; field = valid yield-map pixels; all outside DE4 and DEA

SPLIT

Train: DE4 2019–2020 statistics + fields 2018–2020. Test: DE4 2021 (time), DEA 2019–2020 (space), DEA 2021 (both), fields 2021.

Credits: Miranda, M., Pathak, D., Helber, P., Bischke, B., Najjar, H., Mena, F., ... & Dengel, A. (2026). Yieldsat: A multimodal benchmark dataset for high-resolution crop yield prediction. arXiv preprint arXiv:2604.00940.

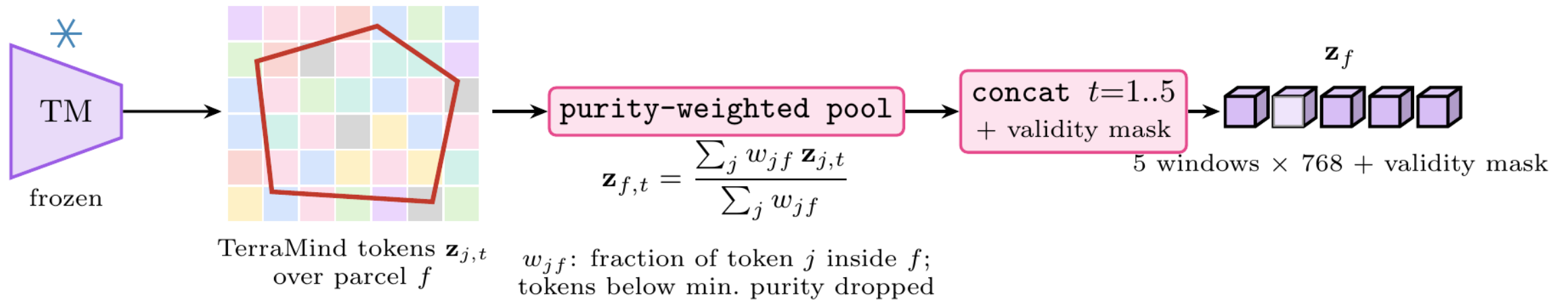
P3. Application of embeddings to crop yield prediction and real-world validation



One YieldSAT field (farm 1, 2018/19), clearest Sentinel-2 acquisition per window.

Credits: Miranda, M., Pathak, D., Helber, P., Bischke, B., Najjar, H., Mena, F., ... & Dengel, A. (2026). Yieldsat: A multimodal benchmark dataset for high-resolution crop yield prediction. arXiv preprint arXiv:2604.00940.

P3 · From imagery to a parcel embedding



1 · Encode

Each window composite is cut into 224-px chips and passed once through frozen TerraMind: 14×14 tokens of 768 values, each token covering 160×160 m.

2 · Pool

Tokens overlapping a parcel are averaged, weighted by the share of the token inside it; tokens less than 25 % inside are dropped.

3 · Stack

The five windows give each parcel a 5×768 vector, with a mask that flags windows lost to cloud.

Computed once on JUPITER and stored; every experiment reuses the same parcel embeddings. The encoder is never retrained.



Thanks!

Fields near Limburgerhof, Germany, Sentinel-2A, 22 Sep 2016. Contains modified Copernicus Sentinel data (2016), processed by ESA, CC BY-SA 3.0 IGO