BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/b

CLOSING THE LOOP FROM SATELLITE EMBEDDINGS TO WATER ALLOCATION: A GEOSPATIAL FOUNDATION MODEL FRAMEWORK FOR CROP SPECIFIC DROUGHT RISK ACCOUNTING IN THE WATER STRESSED IRRIGATED DRYLANDS OF THE ARAL SEA BASIN

Ismoilov Avazbek Abdusamat o‘g‘li
BSc Agrologistics, 2nd year student
International Agriculture University
E-mail: avazbeki733@gmail.com
ORCID: 0009-0009-9659-1514

Abstract

Irrigated agriculture in the Aral Sea Basin consumes roughly ninety percent of all withdrawn water, yet the maps that should inform that allocation and the models that should govern it remain disconnected. Earth observation now produces accurate crop maps, and hydrological tools now produce probabilistic drought risk profiles, but the two communities rarely meet inside a single operational pipeline, and almost never in a setting that is simultaneously transboundary, dominated by fragmented and salinity affected fields, and chronically short of ground truth labels. This paper proposes and argues for a coupled framework that treats the modern geospatial foundation model not as an end in itself but as the connective tissue between the pixel and the water allocation decision. The central idea is to use analysis ready satellite embeddings, in particular the ten meter annual embedding fields produced by foundation models such as AlphaEarth Foundations, together with multitemporal radar and optical features, to derive crop specific classifications and evapotranspiration estimates with very few labels, and then to feed those quantities into a probabilistic water balance and drought risk accounting layer adapted from the D Risk family of decision tools. We specify the components, justify each design choice against the alternatives, and lay out a rigorous validation design with baselines, ablations, cross region transfer tests, and uncertainty propagation. We are deliberately ruthless about failure modes, including phenological information loss in annual embeddings, domain shift across the post Soviet irrigation mosaic, salinity confounding of vegetation signals, the political economy of transboundary data, and the equity consequences of acting on uncertain maps. We close with prioritized, justified recommendations for researchers, water authorities, and the wider Central Asian agricultural statistics enterprise. The contribution is not another crop map and not another water tool, but a defensible architecture for turning self supervised representations of the planet into accountable decisions about a scarce and contested resource.

Keywords: geospatial foundation models; satellite embeddings; crop type mapping; agricultural water accounting; evapotranspiration; probabilistic drought risk; irrigation; Aral Sea Basin; Central Asia; transfer learning; decision support

Introduction

Water, not land and not labor, is the binding constraint on agriculture across most of Central Asia. The Aral Sea Basin, fed principally by the Amu Darya and the Syr Darya, supports irrigation

systems in which agriculture accounts for the overwhelming majority of consumptive water use, and in which a single rotation, cotton followed by winter wheat, still occupies most of the irrigated area in Uzbekistan. The historical record is unambiguous and catastrophic: the diversion of river water to cotton was the principal driver behind the desiccation of the Aral Sea, which has shrunk to a small fraction of its former volume and has produced a new salt desert on the exposed seabed. The pressures are not historical curiosities. Uzbekistan is now routinely listed among the countries expected to face acute water stress within the next fifteen years, climate change is reducing the glacial and snowpack storage that feeds the rivers, soil salinity forces farmers to spend a substantial additional fraction of their water budget on leaching, and new upstream infrastructure such as the Qosh Tepa canal in Afghanistan threatens to remove a large volume of Amu Darya flow before it reaches downstream users. National policy has responded with ambitious water saving targets, including a large planned expansion of drip irrigation and a goal of placing water saving technologies on a substantial share of irrigated land, alongside the construction of an integrated drought management capability built on monitoring and early warning.

These policies require information that is spatially explicit, crop specific, timely, and trusted. A drip irrigation subsidy is only as good as the map that tells the authority which fields grow which crop, how much water those crops actually transpire, and how that demand collides with the volume the river is likely to deliver in a dry year. This is precisely the information that satellite remote sensing has spent five decades learning to produce, and the past three years have transformed what is technically feasible.

The first transformation is methodological. The dominant paradigm in operational crop mapping, supervised classification of multitemporal optical and radar time series with random forests, support vector machines, or convolutional and recurrent networks, has produced excellent results where dense labels and clean phenological time series are available. Within the body of work that informs this paper, time series classification of Sentinel optical data with random forests and phenologically aware features has exceeded ninety percent overall accuracy for several crop types, multisensor fusion of radar and optical indices has pushed cropland delineation accuracy still higher, and semantic segmentation architectures have mapped crop types at the parcel scale. The recurring and unresolved difficulty is that all of these methods are hungry for labels, and they degrade when transferred to a new region, a new year, or a new sensor configuration. The second transformation addresses exactly that weakness. Geospatial foundation models, pretrained by self supervision on enormous unlabeled archives, now offer transferable representations that can be adapted to downstream tasks with a small fraction of the labels previously required. Prithvi EO 2.0, developed by a NASA and IBM led consortium, was trained on millions of global Harmonized Landsat and Sentinel 2 time series samples and demonstrably outperforms both its predecessor and several competing foundation models on standardized benchmarks. TerraMind, a collaboration involving IBM, the European Space Agency, and the Julich Supercomputing Centre, extended the paradigm to any to any multimodal generation across optical, radar, elevation, and other modalities. Most consequentially for our purposes, Google DeepMind released AlphaEarth Foundations, which distills petabytes of multisensor data into compact sixty four dimensional annual embedding fields at ten meter resolution, available globally from 2017 onward, with crop type classification, yield estimation, and evapotranspiration estimation among the tasks already demonstrated on its benchmark suite.

The opportunity, and the gap, become clear when these two literatures are placed beside the water management literature. Decision support tools such as the enhanced D Risk webtool translate crop mix, soil available water capacity, agroclimate, and abstraction constraints into probability

distributions of irrigation deficit and license headroom, giving a farm to catchment view of drought risk. These tools assume their inputs. They assume that someone has told them which crop grows where, how much it needs, and what the supply rules are. Earth observation can now supply two of those three inputs with unprecedented label efficiency and spatial detail, yet the coupling is almost never made, and where it is attempted it is usually in data rich temperate settings rather than in the fragmented, saline, transboundary, label poor irrigation systems where the stakes are highest.

This paper argues that the scientifically interesting and societally urgent problem is not to build a better crop map or a better water tool in isolation, but to close the loop between them, and to do so in the hardest possible setting. We make three contributions. First, we articulate a coupled framework in which foundation model embeddings serve as the shared representational substrate for both crop classification and evapotranspiration retrieval, which then drive a probabilistic water balance and drought risk layer. Second, we specify a concrete and falsifiable validation design for the Aral Sea Basin, with baselines, ablations, transfer experiments, and explicit uncertainty propagation, so that the framework can be tested rather than merely asserted. Third, we conduct an unusually frank analysis of where the approach will break, because a framework that hides its failure modes is worse than useless when the output will be used to allocate a scarce and contested resource among people whose livelihoods depend on it.

Background and state of the art

From hand engineered features to foundation model embeddings. The trajectory of agricultural remote sensing can be read as a steady migration of intelligence from the human analyst into the model. Early operational work relied on vegetation indices such as the normalized difference vegetation index and on the analyst knowledge of crop phenology to select acquisition dates that maximized separability. The machine learning era automated the decision boundary while leaving feature design largely to the analyst, and produced the random forest and support vector machine pipelines that still dominate national crop mapping. The deep learning era automated feature design as well, with convolutional networks for spatial context, recurrent and attention based networks for temporal dynamics, and segmentation architectures for parcel delineation. Throughout, the binding constraint shifted but never disappeared: from analyst time, to labeled samples, to labeled samples that are representative of the deployment distribution.

Foundation models attack the label constraint directly. By pretraining on unlabeled imagery at planetary scale, they learn representations in which the structure relevant to many downstream tasks is already organized, so that a small labeled set suffices to fit a task specific head. Three properties of the current generation matter for irrigated dryland agriculture. The first is label efficiency, which is decisive where field campaigns are expensive and politically sensitive. The second is robustness to missing and corrupted observations, because embeddings learned jointly across optical and radar streams largely absorb cloud gaps, scan artifacts, and sensor differences that otherwise force elaborate gap filling. The third is spatial granularity. The thirty meter native resolution of Harmonized Landsat and Sentinel based models such as Prithvi EO 2.0 is a real liability in landscapes of very small fields, and the agricultural statistics literature has long warned that coarse sensors are unsuitable for the small and very small field categories that dominate much of Asia. The ten meter embedding field of AlphaEarth Foundations is therefore not a cosmetic improvement but a qualitative change in applicability to fragmented irrigation mosaics.

What crop mapping has already established

The studies that anchor this work converge on several robust findings that the proposed framework must respect rather than rediscover. Phenology carries most of the discriminative signal, so any representation that discards the within season trajectory of a crop sacrifices accuracy, a point

that will return as a central tension for annual embeddings. Multisensor fusion of radar and optical data improves both accuracy and resilience, with radar supplying structure and all weather observation while optical supplies biochemical and pigment information. Seasonal structure must be respected explicitly, as in the separate treatment of the winter spring and summer autumn cropping cycles that organizes cropland extraction in the South Asian work and that maps directly onto the Central Asian winter wheat and cotton calendar. Finally, validation against independent field survey and official statistics, not merely against a held out split of the same labels, is what separates a credible operational claim from a benchmark number.

What the water and statistics communities contribute. The D Risk family of tools demonstrates that drought risk in irrigated agriculture is fundamentally probabilistic and multiscale. It is probabilistic because the relevant question is not whether a deficit will occur but with what probability and what magnitude, given an ensemble of plausible climate and flow realizations. It is multiscale because a deficit that is unmanageable at the individual farm can be neutralized at the catchment level through coordination and sharing, a result with obvious resonance in canal command areas where water is allocated collectively. The agricultural statistics literature contributes the discipline of sampling frames and area estimation, the insistence on confusion matrices computed against independent ground data, and the sober mapping between field size distributions and sensor suitability that justifies the move to ten meter representations in Central Asia. Evapotranspiration retrieval from fused Sentinel and Landsat data provides the physical bridge between a crop map and a water volume, since it is consumptive use, not planted area alone, that determines the call on the river.

The unoccupied position. Placing these literatures side by side reveals an unoccupied and valuable position. Foundation models have largely been evaluated as crop mappers, stopping at the classified map. Water tools have largely assumed their crop and demand inputs. Neither community has, to our knowledge, built and validated an end to end pipeline in which self supervised embeddings simultaneously produce the crop map and the consumptive use estimate that drive a probabilistic, multiscale drought risk account in a transboundary, fragmented, saline, label poor basin. That position is what this paper proposes to occupy.

The proposed framework

The framework consists of four coupled layers. We describe each layer, the rationale for its design, and the alternatives we reject and why.

Layer one: a shared embedding substrate. The first layer converts raw multisensor observation into an analysis ready representation that both downstream tasks share. We propose using the ten meter annual satellite embedding field as the spatial backbone, augmented with explicitly multitemporal radar and optical features within the growing season. The rationale is twofold. Using a single shared substrate for both classification and consumptive use estimation enforces physical and statistical consistency between the crop map and the water estimate, so that the two never contradict each other at the pixel. Using the ten meter embedding rather than a thirty meter foundation model is dictated by the field size distribution of Central Asian irrigation, where coarse pixels straddle multiple parcels and destroy the very boundaries that water accounting depends on.

The critical design tension is temporal. Annual embeddings summarize a year, and the discriminative power of crop classification lives in the within season trajectory. Cotton and winter wheat are separable largely because their green up, peak, and senescence occur at different times. Compressing a year into a static embedding risks discarding exactly this signal. We therefore reject the naive option of using the annual embedding alone. The framework instead treats the annual embedding as a strong, cloud robust spatial prior and concatenates it with a compact set of

intraseasonal features, for example radar backscatter and optical index statistics computed within the winter spring and summer autumn windows that define the local calendar. This preserves phenology where it matters while inheriting the label efficiency and gap robustness of the embedding. An alternative we hold in reserve is to use a multitemporal foundation model that ingests the time series natively, accepting its coarser resolution and compensating with subpixel methods; we treat this as a comparison arm rather than the default, because the resolution penalty is severe in this landscape.

Layer two: crop specific classification with minimal labels. The second layer fits a lightweight classification head on the shared substrate, targeting the crop classes that matter for water governance: cotton, winter wheat, rice where present, orchards and vineyards, fodder and the increasingly important diversification crops, fallow, and abandoned or salinized land. The rationale for a lightweight head, rather than full finetuning of a large backbone, is both practical and statistical. It is practical because the computational and data resources available to a national water authority are limited. It is statistical because, in the label poor regime, full finetuning of a large model invites overfitting and the amplification of class imbalance, a failure already documented when a large foundation model was fully finetuned on fragmented smallholder data. The framework therefore defaults to frozen embeddings with a small trained head and a calibrated treatment of class imbalance, and uses full finetuning only as an ablation to quantify what, if anything, it buys.

Layer three: from crop type to consumptive water use. The third layer converts the classified, parcel resolved map into a spatially explicit estimate of crop water demand and actual consumptive use. Two complementary paths are specified. The first is a crop coefficient path, in which the crop class and its phenological stage index a reference evapotranspiration to yield a theoretical demand, the classical and transparent approach that water balance tools already use. The second is a direct retrieval path, in which actual evapotranspiration is estimated from fused thermal, optical, and radar observation, capturing the effect of water stress, salinity, and deficit irrigation that the coefficient path cannot see. The rationale for carrying both is that their disagreement is itself informative. Where retrieved actual evapotranspiration falls below the theoretical demand, the gap is a direct, physically grounded signal of water stress, which is far more actionable than either quantity alone. This is a deliberate departure from tools that model demand without observing supply side stress.

Layer four: probabilistic drought risk and allocation accounting. The fourth layer is the one most often missing from remote sensing studies and the one that makes the output a decision rather than a description. It adapts the probabilistic water balance logic of the D Risk family to the Central Asian context. The adaptation is substantial and is itself a contribution, because the institutional reality differs from the licensed abstraction regime for which those tools were built. In place of individual surface and groundwater licenses with fixed volumetric limits, the framework represents canal command areas with collectively allocated and seasonally negotiated deliveries, transboundary flow that is uncertain and contested, and salinity driven leaching demand that must be added to crop demand. Driven by an ensemble of plausible seasonal flow realizations, the layer produces, for each command area and each scale from field to province, a probability distribution of irrigation deficit, the probability and expected magnitude of failing to meet demand including leaching, and the marginal value of coordination, that is, how much deficit could be removed by reallocating water across fields and across cooperating users. The output is therefore not a single number but a risk surface that a water authority can act on, including the identification of which command areas are most exposed in a developing dry year, early enough to matter.

Why coupling beats a pipeline of independent steps. A reasonable critic will ask why these four layers must be co designed rather than chained as independent off the shelf components. The answer is uncertainty. Each layer has error, and a chain that passes only point estimates forward produces a final risk number with no credible confidence. Co design allows uncertainty to be propagated coherently: classification probabilities become soft crop fractions rather than hard labels, evapotranspiration retrievals carry error bars, and the flow ensemble carries hydrological uncertainty, so that the final deficit distribution honestly reflects all three sources. A water authority that is told a command area faces a forty percent probability of a severe deficit, plus or minus a stated and decomposed uncertainty, can govern; one that is handed a deterministic map of unknown reliability cannot, and may do harm by acting on spurious precision.

Proposed implementation and study design

Study area. The proposed demonstration centers on the irrigated zones of the Amu Darya and Syr Darya within Uzbekistan, with Karakalpakstan and Khorezm representing the downstream, salinity stressed, supply insecure end of the Amu Darya, and the Syrdarya and Fergana command areas representing intensively managed systems. These zones span the full range of difficulty: very small to medium fields, the cotton and winter wheat rotation with growing diversification, severe and spatially variable soil salinity, and acute sensitivity to upstream decisions and to dry years. They are also where national water saving policy is being implemented, so that an operational risk product would have an immediate institutional home.

Data. The spatial backbone is the ten meter annual satellite embedding field. The intraseasonal augmentation uses radar backscatter and dual polarization descriptors for structure and all weather observation, and optical surface reflectance and vegetation, water, and red edge indices for biochemical signal, composited within the two cropping windows. The consumptive use layer adds thermal and meteorological inputs for evapotranspiration retrieval. Hydrological inputs comprise river flow records and an ensemble of seasonal flow scenarios that represent both natural variability and plausible upstream abstraction, including the contingency of large diversions such as the Qosh Tepa canal. Reference data combine the sparse available field surveys, official agricultural and water statistics at district and province level, and a purpose built ground campaign, with the explicit understanding that labels will be scarce, which is the regime the framework is designed for.

Label efficient training and the role of statistics. Because labels are scarce and field access is constrained, the training strategy leans on three devices. It uses the embedding substrate to minimize the number of labels required for a given accuracy. It uses official subnational statistics as an aggregate constraint, requiring that the mapped crop areas, when summed to district and province, are consistent with reported areas, a soft supervision signal that the agricultural statistics tradition has long exploited through area frame and small area estimation methods. And it uses transfer from better labeled analog regions as an initialization, to be corrected by the few local labels available. The interplay between pixel level mapping and aggregate statistics is not a nuisance to be tolerated but a resource to be designed in.

Adapting the risk layer to canal command institutions. The risk layer must be reformulated for an institutional setting of collective canal allocation and transboundary uncertainty. We propose to represent each command area as the unit of allocation, to model delivery as a seasonally negotiated share of an uncertain basin flow rather than a fixed license, to add salinity driven leaching demand to crop demand using the salinity signal extracted from the embeddings and optical indices, and to compute the coordination value across fields within a command area and across command areas within a province. This reformulation is necessary because importing the licensed abstraction logic

unchanged would misrepresent the system and produce risk numbers that no Central Asian water manager would recognize or trust.

Validation design

A framework earns the right to be believed only through a validation design that could prove it wrong. We specify the following.

Tasks, metrics, and baselines. Crop classification is evaluated by overall accuracy, per class producer and user accuracy, and the area under the precision recall curve for the minority diversification classes that matter most for policy and that conventional accuracy hides. Crucially, accuracy is reported both per pixel and per parcel, and area estimates are reported with confidence intervals and compared against independent official statistics, following the discipline of the agricultural statistics tradition rather than the convenience of a held out split. Evapotranspiration is evaluated against any available flux or lysimeter reference and against basin scale water balance closure. The risk layer is evaluated by the calibration of its probabilistic forecasts, that is, whether events assigned a forty percent probability occur about forty percent of the time, assessed with reliability diagrams and proper scoring rules, and by hindcasting known historical dry years to test whether the system would have flagged the command areas that actually suffered.

The baselines are deliberately strong, because beating a weak baseline proves nothing. They include a supervised random forest on multitemporal optical and radar features, the standard operational approach; a deep temporal network trained on the same time series; and a thirty meter multitemporal foundation model. The framework must demonstrate advantage against these, not against a strawman.

Ablations. The ablations isolate the contribution of each design decision. We remove the intraseasonal augmentation to quantify the phenological information lost by the annual embedding alone, the central tension of the framework. We compare frozen embeddings with full finetuning to test the overfitting hypothesis in the label poor regime. We remove the actual evapotranspiration path to measure how much the supply side stress signal adds over a demand only model. We remove the statistical aggregate constraint to measure its regularizing value. We remove uncertainty propagation to demonstrate, by its absence, the danger of deterministic outputs.

Transfer and generalization. Because the operational promise is generalization, we test it directly. We train in better labeled command areas and evaluate, without local labels, in held out command areas and in held out years, including a year not seen in training. We report the accuracy decay under transfer, since the honest measure of a foundation model is not its peak in distribution accuracy but its graceful degradation out of distribution. A method that scores highly in sample but collapses across a canal boundary or across a dry year is not fit for the purpose claimed here.

Success criteria, stated in advance. To avoid moving the goalposts, success is defined before results are seen. The framework succeeds if, with an order of magnitude fewer labels than the supervised baseline, it matches or exceeds that baseline on per parcel crop accuracy and on area estimates against official statistics; if its risk forecasts are well calibrated on historical dry years; and if it retains usable accuracy under cross command area and cross year transfer. Failing any of these is a result worth reporting, not a result to be buried.

Anticipated results, tradeoffs, and honest uncertainty

We state expectations as hypotheses with reasoning, not as findings, and we attach our confidence to each.

We expect, with moderate to high confidence, that the embedding substrate will substantially reduce the labels required to reach a given crop accuracy relative to the supervised baseline, because label efficiency is the property foundation models most reliably deliver and the property

most relevant to this setting. We expect, with moderate confidence, that the annual embedding alone will underperform on the separation of phenologically similar classes, and that the intraseasonal augmentation will recover most of the gap; if it does not, the framework central design choice is wrong and must be revised toward a natively multitemporal backbone. We expect, with lower confidence, that the actual evapotranspiration path will add meaningfully over the demand only model specifically in salinity stressed and deficit irrigated areas, where the stress signal is strongest and where the policy need is greatest. We are genuinely uncertain whether the risk layer will be well calibrated, because calibration depends on the seasonal flow ensemble, which is the least controlled input and the one most exposed to transboundary unpredictability.

The tradeoffs are real and must be named. Ten meter resolution buys applicability to small fields at the cost of a heavier data and compute burden than thirty meter products. Embeddings buy label efficiency and gap robustness at the cost of interpretability, since a sixty four dimensional learned vector does not explain itself to a water manager, and we therefore treat interpretability as a first class requirement rather than an afterthought, pairing every risk output with the physical quantities, crop type, consumptive use, and supply shortfall, that a human can audit. Annual embeddings buy cloud robustness at the cost of phenological detail, the tension already discussed. Probabilistic outputs buy honesty at the cost of simplicity, since a distribution is harder to communicate than a number, which places a real burden on the design of the decision interface.

Risks, failure modes, and limitations

This section is intentionally adversarial toward our own proposal. Domain shift is the most likely cause of operational failure. The post Soviet irrigation mosaic is heterogeneous in field geometry, management, and data availability, and a model that learns the texture of one command area may misread another. Our transfer experiments are designed to surface this, but surfacing it is not solving it, and a negative transfer result would constrain the claim to within region deployment with periodic local retraining.

Label scarcity is not merely a quantity problem but a representativeness problem. If the few available labels oversample accessible, large, well managed fields near roads, the model will be biased against the small, remote, salinized fields where water stress is worst, and the bias will be invisible in a conventional accuracy figure. This is why we insist on validation against independent aggregate statistics and on minority class metrics.

Phenological information loss in the annual embedding could prove more damaging than anticipated, particularly for distinguishing the diversification crops that policy most wants to track. If the intraseasonal augmentation does not recover it, the framework must change.

Salinity is a confounder, not only a class. Salt stress depresses vegetation indices in ways that can mimic water stress or alter the spectral signature of a healthy crop, so the consumptive use and stress signals are entangled with salinity. Disentangling them is a research problem in its own right, and we flag it as a known weakness rather than pretending it is solved.

The hydrological input is the weakest link. Transboundary flow is uncertain, contested, and subject to upstream decisions that are political rather than hydrological, including new diversions. A risk layer is only as credible as its flow ensemble, and we are explicit that the ensemble, not the imagery, is likely to dominate the final uncertainty.

The political economy of data is a limitation that no algorithm addresses. Water in this basin is allocated across sovereign states with divergent interests, ground truth is sensitive, and a transparent, accurate map of who uses how much water can be politically inconvenient. A technically excellent product that no authority will adopt or share has not solved the problem it set out to solve.

Finally, there is an equity hazard intrinsic to acting on uncertain maps. A risk product that misclassifies a smallholder crop or overestimates a command area consumptive use can justify the withdrawal of water from people who can least absorb the loss. The appropriate response is not to withhold the technology but to surface uncertainty honestly, to keep a human authority in the loop, and to treat the map as advisory evidence rather than automatic instruction.

Policy and governance implications

The framework is built to serve decisions, and its implications are concrete. For the implementation of water saving policy, a parcel resolved, crop specific consumptive use map allows subsidies for drip irrigation and for crop diversification to be targeted at the fields and command areas where the water saving per dollar is greatest, rather than distributed uniformly. For drought management, the probabilistic risk surface is precisely the monitoring and early warning input that an integrated drought management system requires, capable of flagging exposed command areas early in a developing dry year. For agricultural statistics, the same pipeline that drives the risk layer can produce timely, spatially explicit crop area and water use estimates that complement and cross check official statistics, advancing the long standing goal of using remote sensing to make agricultural statistics more sustainable and more granular. For transboundary cooperation, a shared, transparent, basin wide accounting of consumptive use, however politically delicate, is a precondition for any durable agreement on flow sharing, and an objective evidence base can shift negotiation from assertion toward measurement.

Recommendations and next steps

The following recommendations are ordered by priority, each with its justification, and each is a concrete action rather than an aspiration.

First, build and release the intraseasonal augmentation experiment before anything else. The entire framework rests on the claim that annual embeddings plus a small set of seasonal features recover the phenological signal. This claim is cheap to test and decisive in its consequences, so it should be the first result produced. If it fails, the backbone choice must change immediately, and every later effort would otherwise be wasted.

Second, treat the seasonal flow ensemble as a first class research object, not as a fixed input. The analysis in this paper makes clear that hydrological uncertainty, including upstream diversion, will likely dominate the final risk uncertainty. Investing in a credible, well characterized flow ensemble, including explicit scenarios for new infrastructure, will do more for the trustworthiness of the output than any further gain in imagery accuracy.

Third, co design the validation with the institution that will use the output. The success criteria stated here are technical, but adoption depends on institutional criteria, and the risk surface must be expressed in the command area units and decision rhythms that water managers actually use. The fastest path to irrelevance is a beautiful product calibrated to the wrong unit of decision.

Fourth, make uncertainty and interpretability mandatory deliverables, not optional extras. Because the output will be used to allocate scarce water among vulnerable people, every risk number must arrive with a decomposed uncertainty and with the auditable physical quantities behind it. This is both an ethical requirement and a practical one, since unexplained outputs will not be trusted and should not be.

Fifth, design for label scarcity as the permanent condition, not as a temporary inconvenience. The recommendation is to combine embedding based label efficiency, aggregate statistical constraints, and transfer from analog regions into a standing methodology, and to budget explicitly for a small, carefully stratified ground campaign that deliberately oversamples the small, remote, and salinized fields that convenience sampling would miss.

Sixth, pursue salinity disentanglement as a dedicated subproject. Because salinity confounds the water stress signal that is central to the policy value of the framework, separating the two deserves its own focused effort, likely combining the embedding representation with ancillary soil and groundwater data.

Seventh, plan for transboundary data governance from the outset. The technical and the political cannot be separated in this basin. Engaging the relevant basin organizations early, and designing the product to support rather than threaten cooperative allocation, is a precondition for impact and should be treated as part of the research, not as a downstream afterthought.

Conclusion

The instruments now exist to see every irrigated field in the Aral Sea Basin at ten meters, to infer its crop with a handful of labels, to estimate how much water it consumes, and to weigh that demand against an uncertain river. What has been missing is not capability but connection. This paper has argued that the worthwhile and urgent task is to close the loop from the self supervised representation of the planet to the accountable allocation of its scarcest resource, and to do so in the hardest setting rather than the easiest. We have specified a coupled framework, justified its design against the alternatives, laid out a validation that could refute it, and named the failure modes that a less careful treatment would hide. The framework is not a finished result and does not pretend to be. It is an architecture and a research agenda, offered in the conviction that the value of geospatial foundation models will ultimately be measured not by their benchmark scores but by whether they help a water authority in Karakalpakstan make a better decision in a dry year, and by whether they do so without quietly transferring the cost of their errors onto the people least able to bear it.

References

1. Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C. (2010). A generalized regression based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sensing of Environment, 114(6), 1312 to 1323.
2. Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., Kohli, P. (2025). AlphaEarth Foundations: a geospatial embedding model and the Satellite Embedding dataset. Google and Google DeepMind. Earth Engine dataset GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL.
3. Chengot, R., Knox, J. W., Coxon, G., Cojocaru, G., Holman, I. P. (2023). An enhanced version of the D Risk decision support webtool for multiscale management of water abstraction and drought risks in irrigated agriculture. Computers and Electronics in Agriculture, 204, 107516.
4. Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., et al. (2019). Near real time agriculture monitoring at national scale at parcel resolution: performance assessment of the Sen2 Agri automated system in various cropping systems around the world. Remote Sensing of Environment, 221, 551 to 568.
5. Dimou, S., Stamou, A., Kaimaris, D., Patias, P. (2021). Mapping and monitoring of agricultural land with time series satellite data. EARSeL Symposium proceedings.
6. Gallego, J. (2017). Use of remote sensing for the design of sampling frames. In Handbook on Remote Sensing for Agricultural Statistics. Global Strategy to improve Agricultural and Rural Statistics (GSARS).
7. Guzinski, R., Nieto, H., Ramo Sanchez, R., Sanchez, J. M., Jomaa, I., Zitouna Chebbi, R., Roupsard, O., Lopez Urrea, R. (2023). Improving field scale crop actual evapotranspiration

monitoring with Sentinel 3, Sentinel 2, and Landsat data fusion. International Journal of Applied Earth Observation and Geoinformation, 125, 103587.

8. Jakubik, J., Roy, S., Phillips, C. E., Fraccaro, P., Godwin, D., Zadrozny, B., Szwarcman, D., et al. (2024). Prithvi EO 2.0: a versatile multitemporal foundation model for Earth observation applications. arXiv:2412.02732.
9. Raza, D., Shu, H., Nazeer, M., Aslam, H., Mirza, S., Xiao, X., Sardar, A., Aeman, H. (2024). Improved method for cropland extraction of seasonal crops from multisensor satellite data. International Journal of Remote Sensing.
10. TerraMind (2025). A multimodal any to any geospatial foundation model. IBM, European Space Agency, and Julich Supercomputing Centre.
11. Wang, Y., Braham, N. A. A., Xiong, Z., Liu, C., Albrecht, C. M., Zhu, X. X. (2023). SSL4EO S12: a large scale multimodal multitemporal dataset for self supervised learning in Earth observation.
12. Waldner, F., et al. (2015). Field size and the suitability of satellite sensors for agricultural monitoring, as discussed within the GEOGLAM framework.
13. International Water Management Institute (2025). Uzbekistan country profile: water scarcity and agricultural water use.
14. Nurbekov, A., et al. (2025). Water scarcity, crop diversification, and the cotton and winter wheat rotation in Uzbekistan. Frontiers in Sustainable Food Systems.
15. Carnegie Endowment for International Peace (2025). The Qosh Tepa canal and the prospect of a Central Asian water crisis.