Hybrid Probabilistic Deep Learning for Drought Monitoring

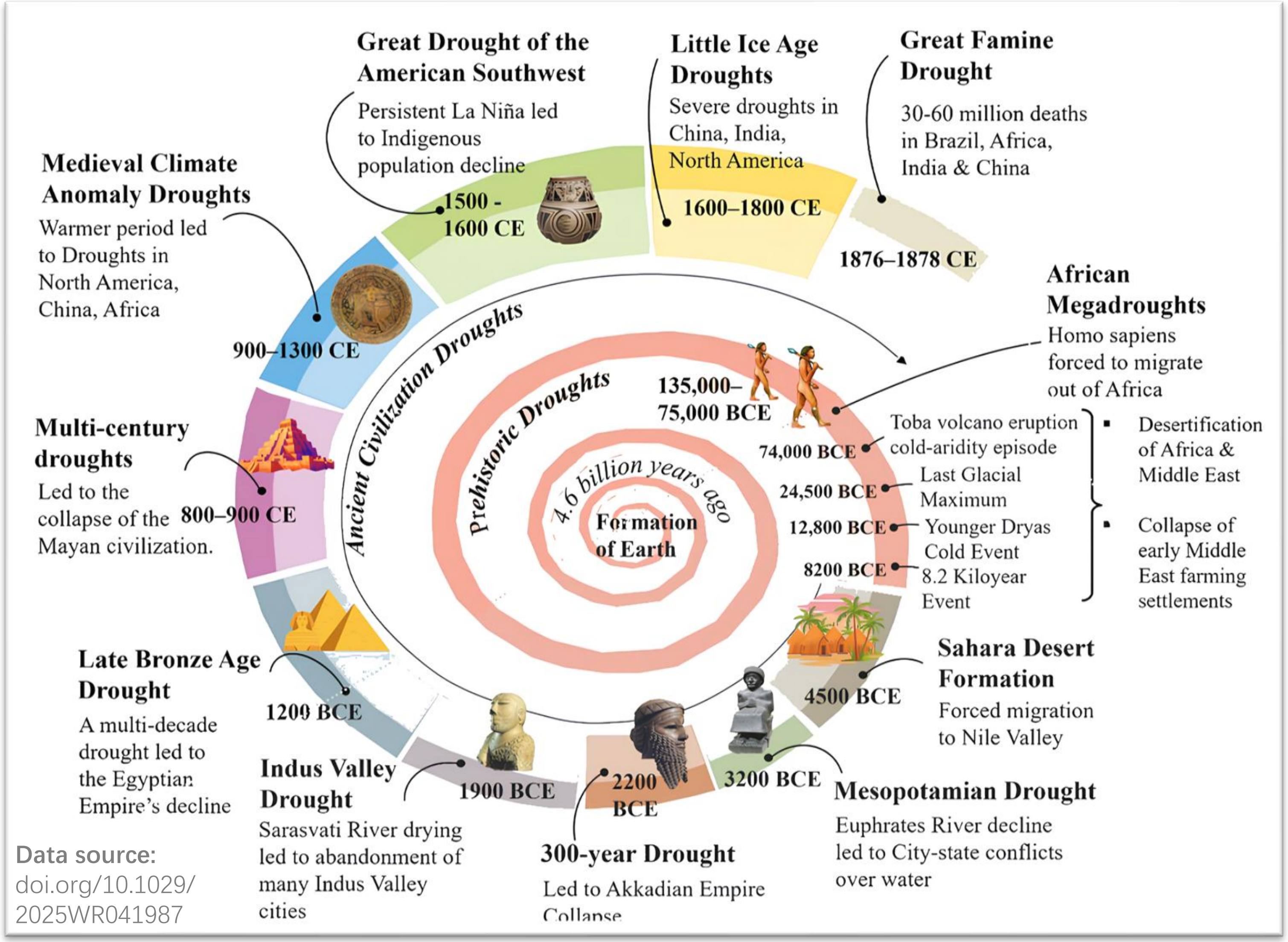
PhD candidate: Mengxue Zhang
Supervisors: Gustau Camps-Valls,
Miguel-Ángel Fernández-Torres

7th July 2026

Doctoral Program in Remote Sensing

- **Introduction**
- **Extreme Drought Event Detection**
- **Reliable Probabilistic Drought Prediction**
- **Long-term Soil Moisture Drought Forecasts**
- **Conclusion**

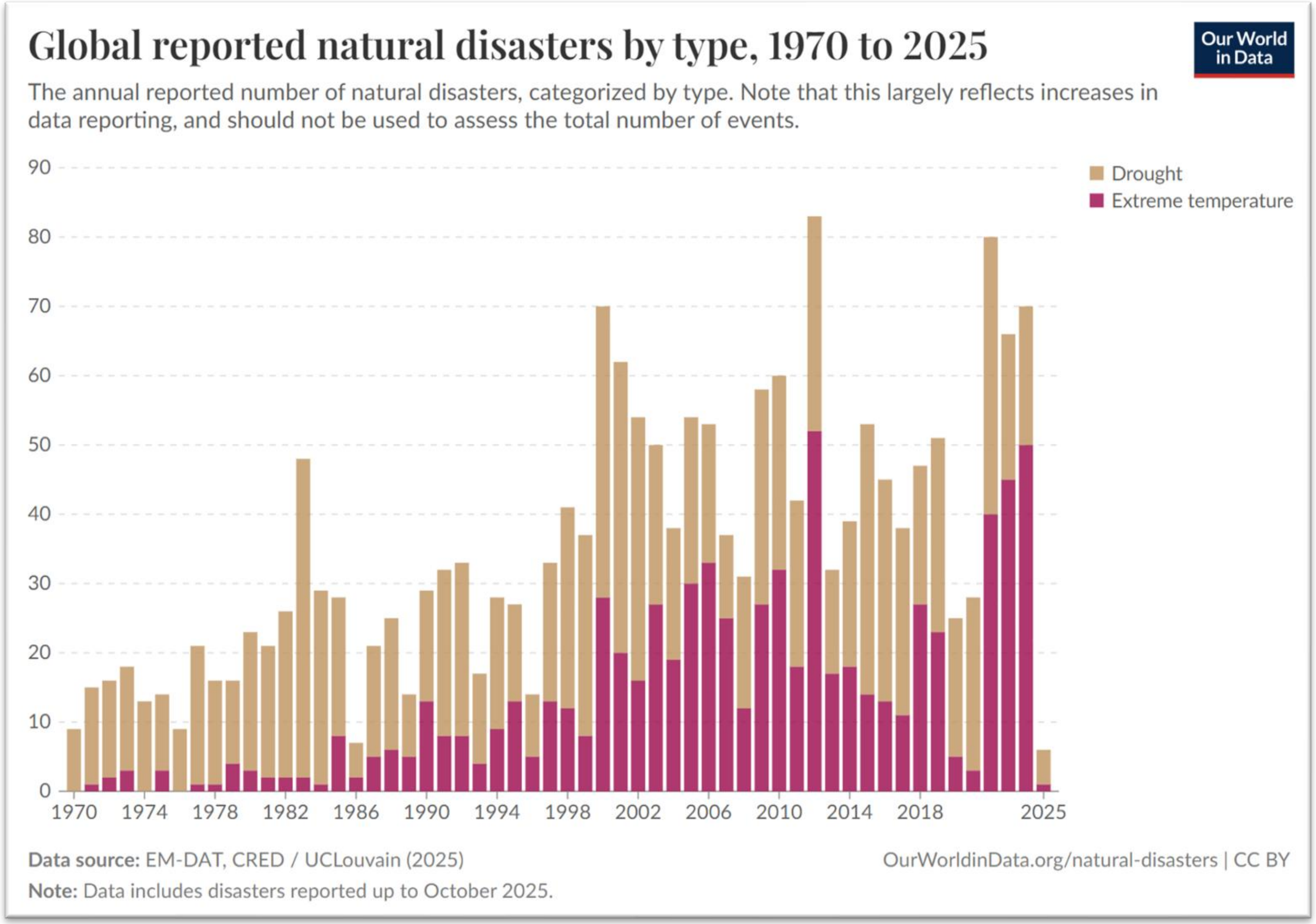
Why Focus on Drought - Impact Spans the Ages



Data source: doi.org/10.1029/2025WR041987

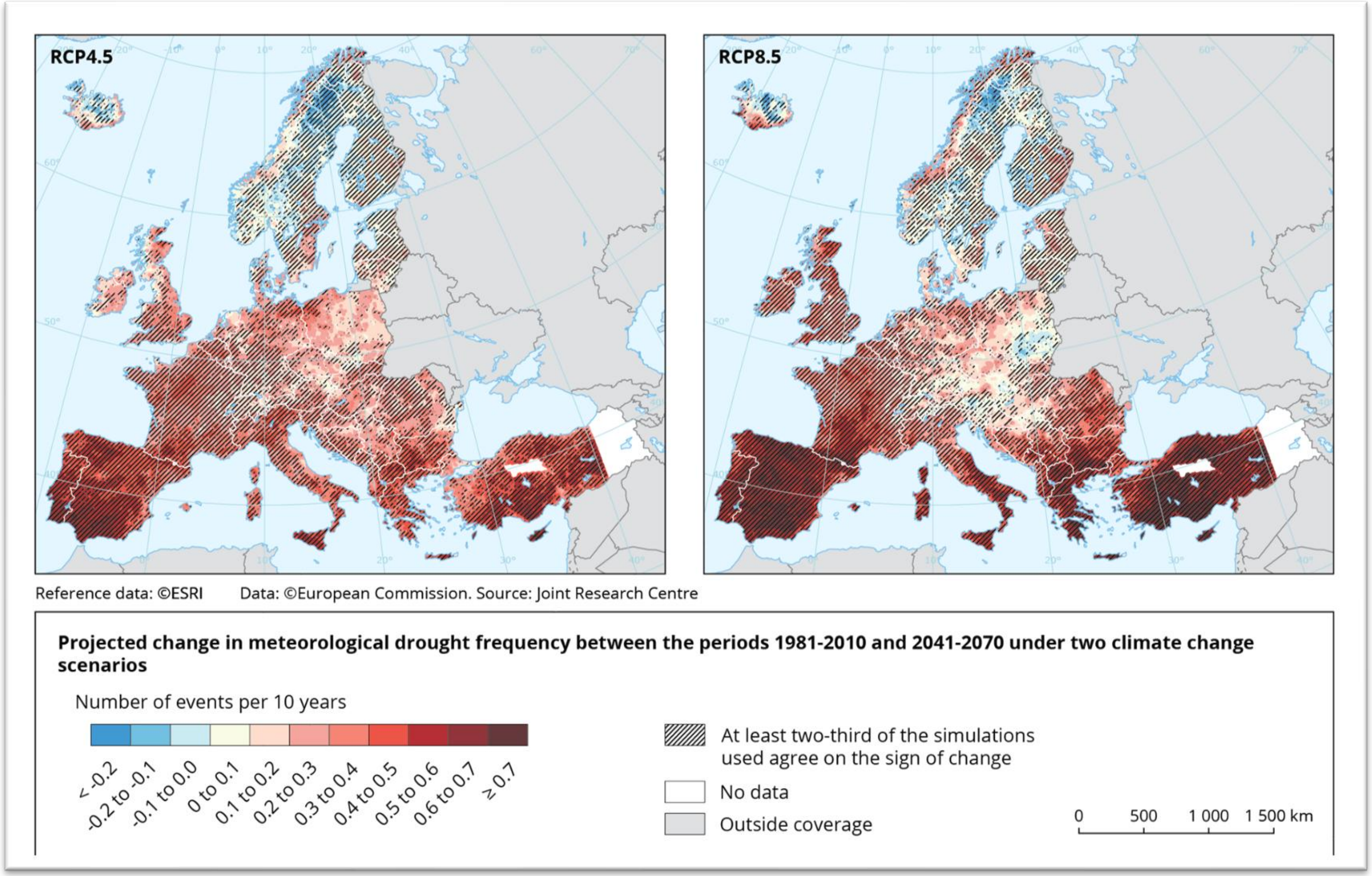
The historical drought

Why Focus on Drought - Impact Spans the Ages



The current drought

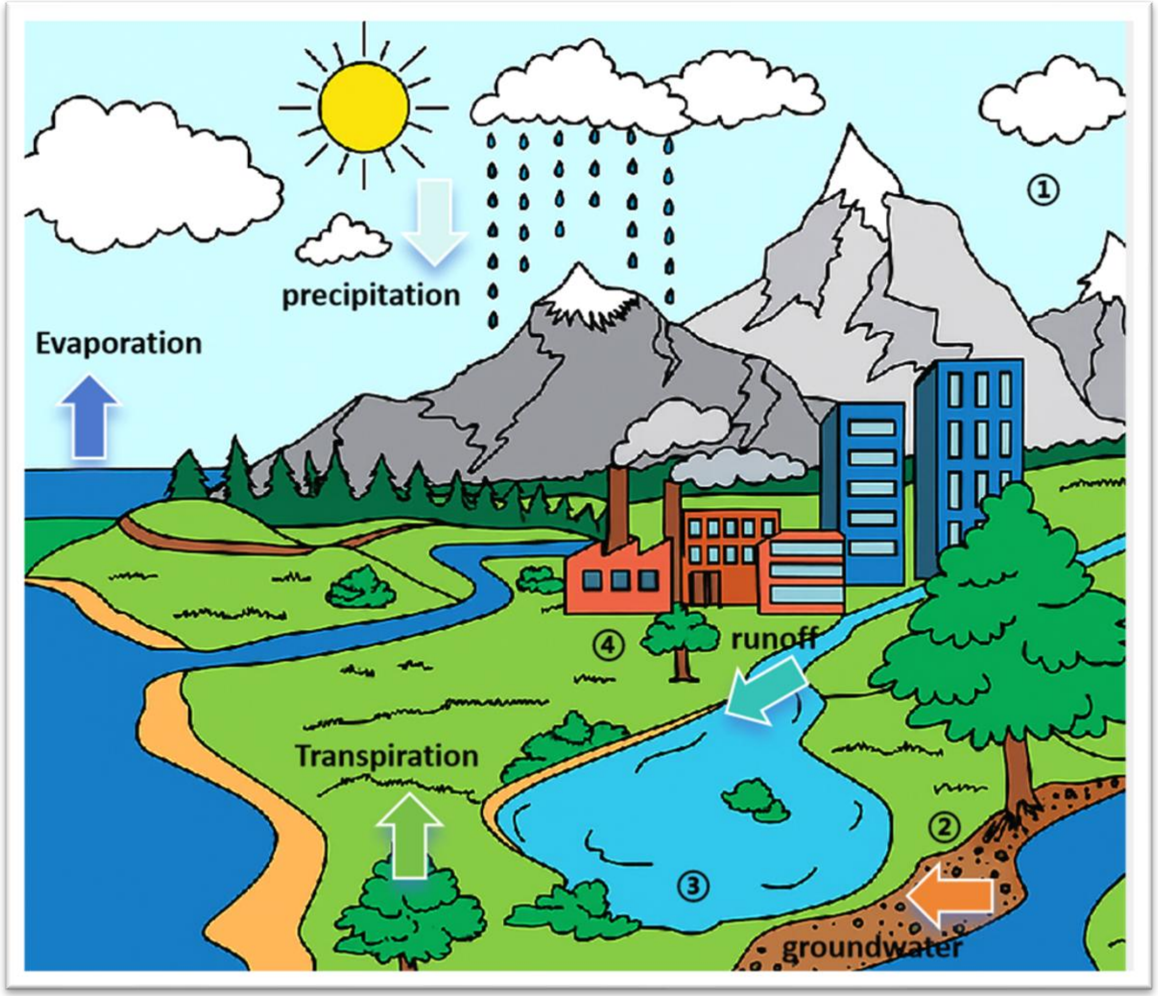
Why Focus on Drought - Impact Spans the Ages



The future drought

Why Focus on Drought - Multifaceted and Dynamic

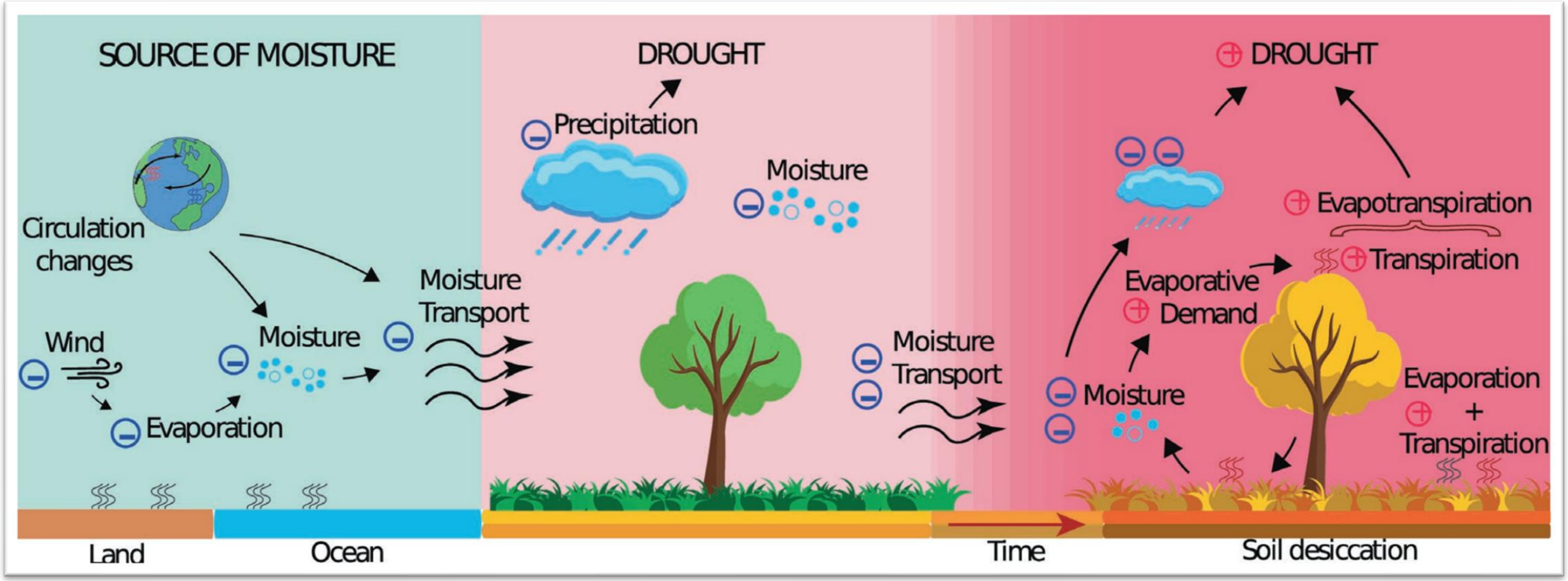
The type of drought



① Meteorological drought
② Agricultural drought
③ Hydrological drought
④ Socioeconomic drought

Data source:
https://thehomeschoolscientist.com/learning-water-cycle-experiment/

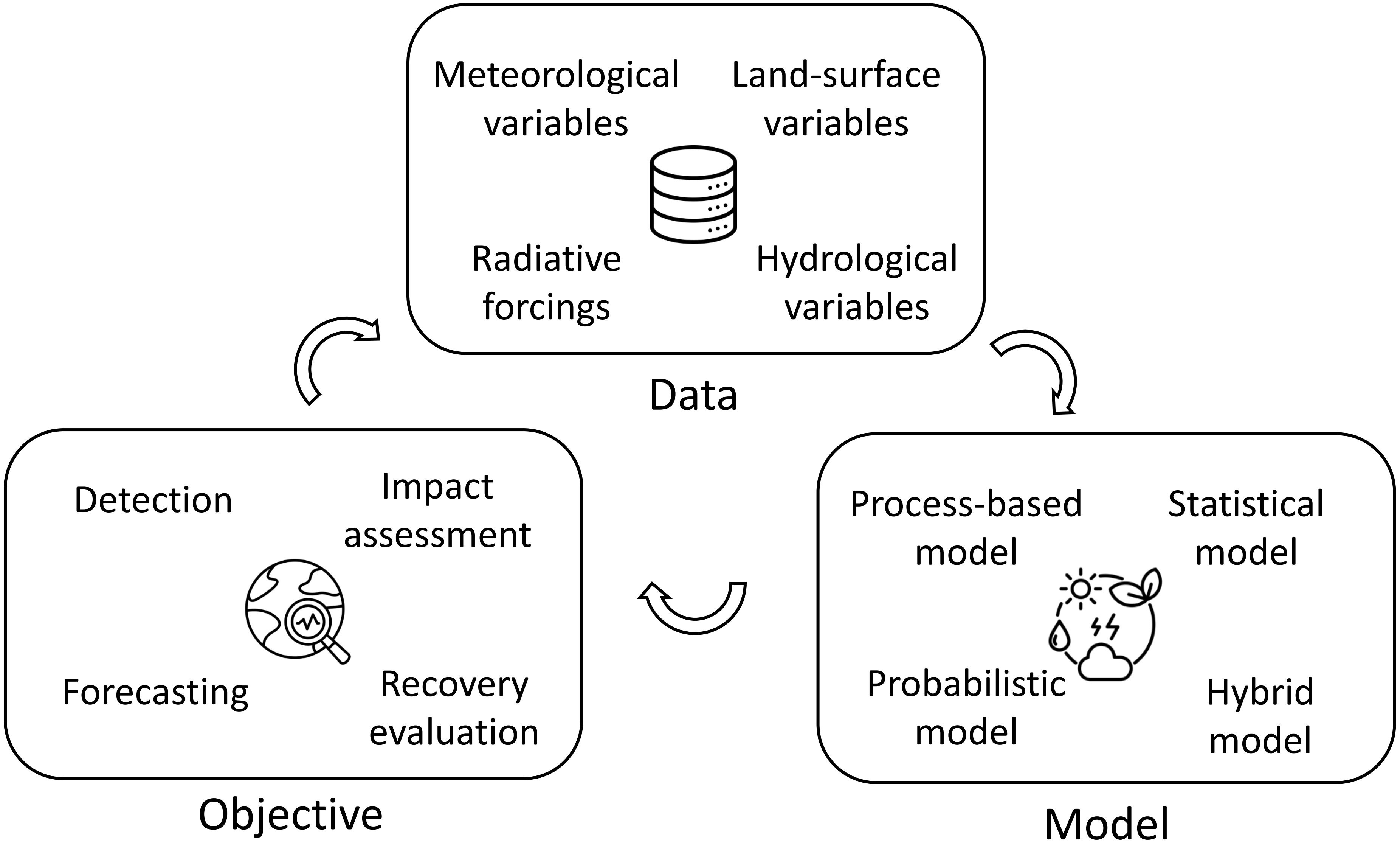
Why Focus on Drought - Multifaceted and Dynamic



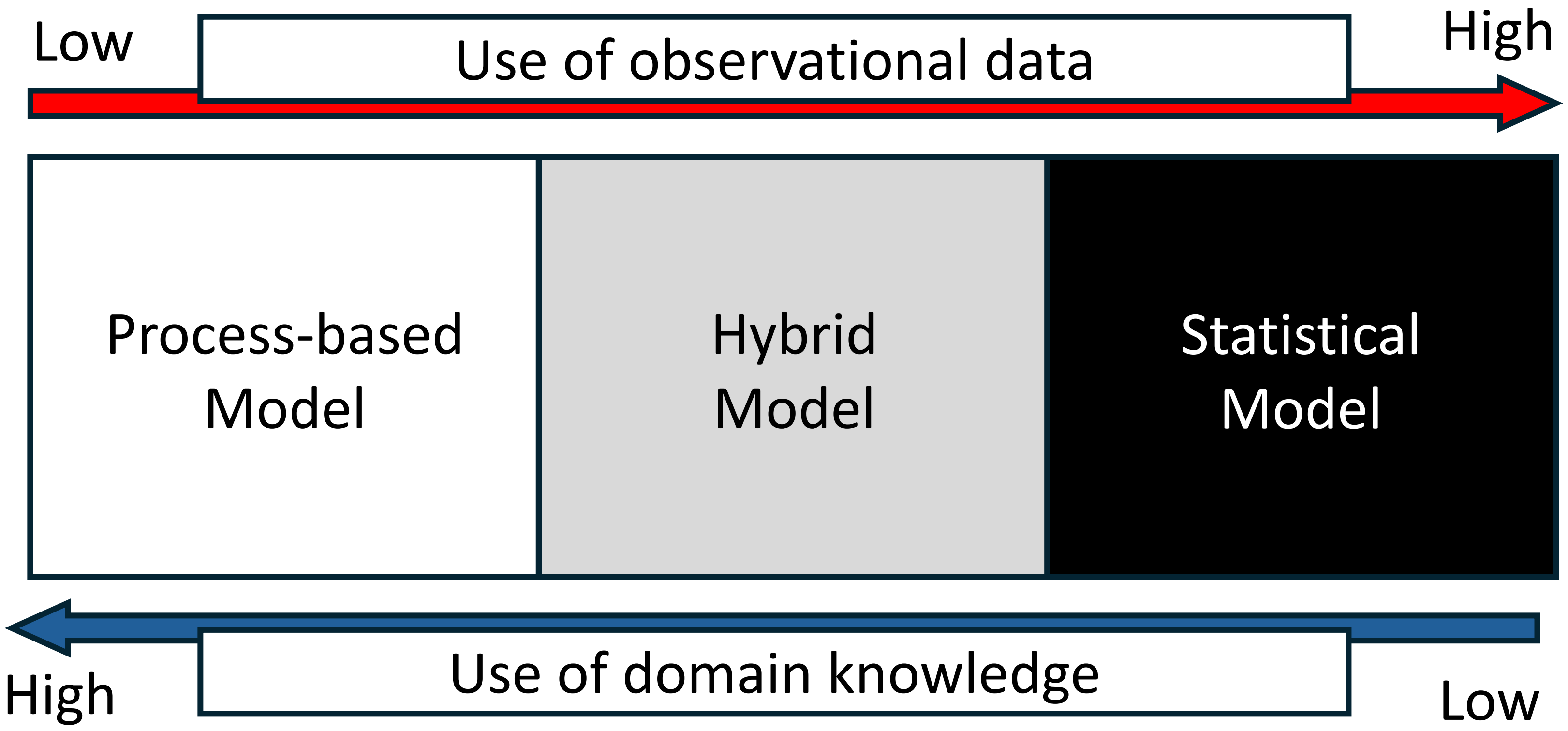
Data source: doi.org/10.1002/wcc.70030

The maintenance and intensification of drought

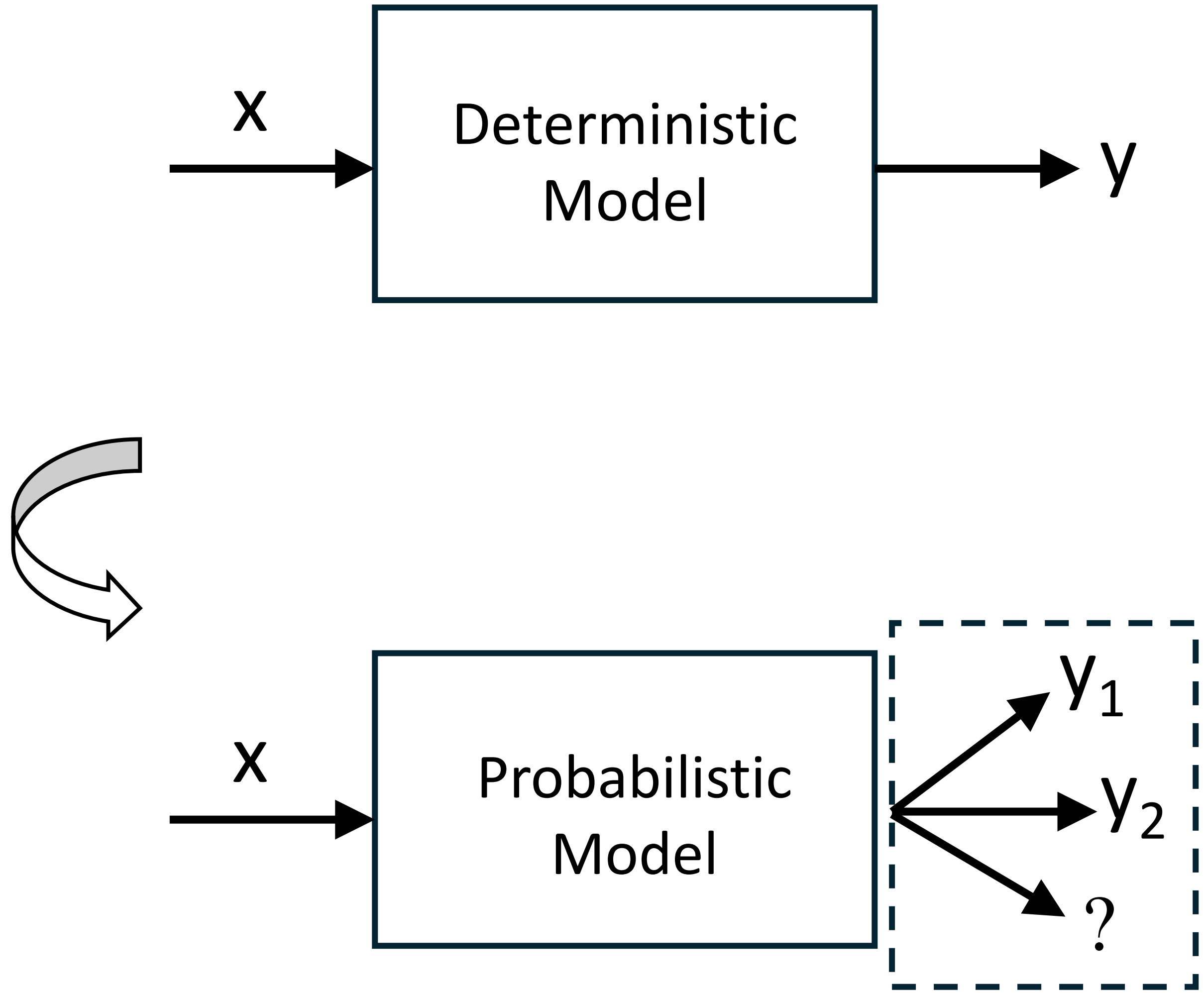
Advances and Limitations of Methods - Overview



Advances and Limitations of Methods – Hybrid Modeling



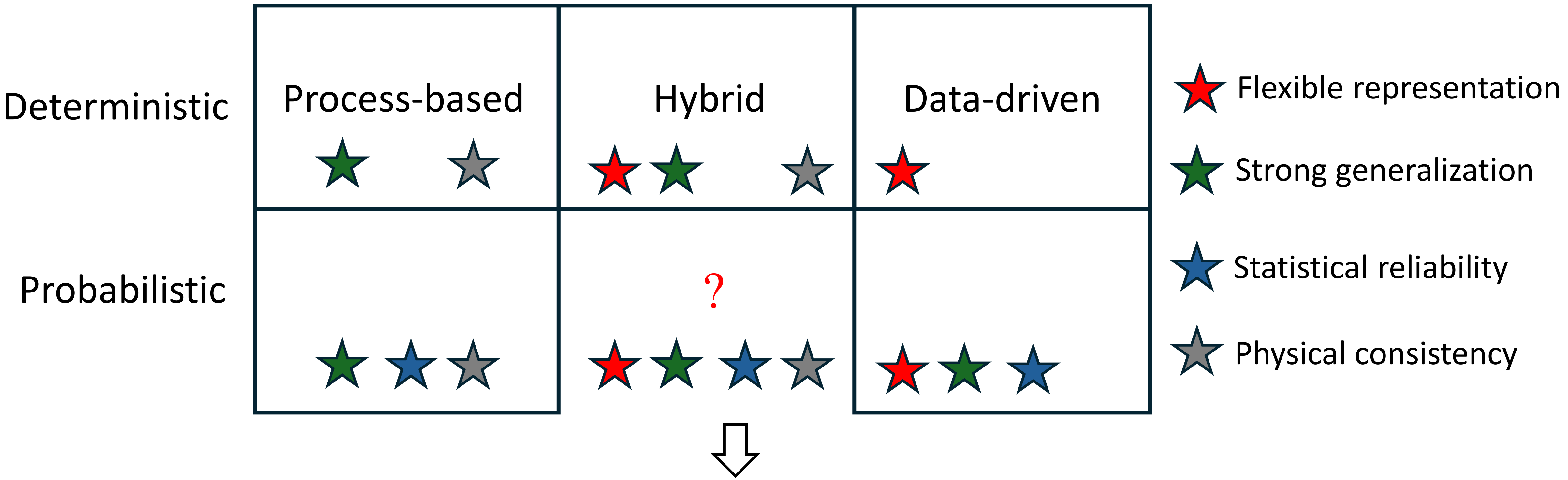
Advances and Limitations of Methods – Probabilistic Modeling



Research Gap and Problem Statement

	Process-based	Hybrid	Statistical
Deterministic	Land surface model Hydrological model	Physics-informed NN Differential ESM	Drought index Random forest, NN
Probabilistic	Ensemble simulation Stochastic parameter		Gaussian process Bayesian NN

Research Gap and Problem Statement



Key Question:

How can we integrate process knowledge, deep learning, and probabilistic methods into a unified framework for accurate, reliable, physically consistent drought modeling?

Research Objectives

Aim:

Develop hybrid probabilistic deep learning frameworks that jointly enhance reliability and consistency for drought applications

Study #1
Drought Event Detection

Domain Knowledge + Probabilistic DL

Study #2
Probabilistic Drought Prediction

Calibration + Uncertainty Quantification

Study #3
Long-term SM Drought Forecasts

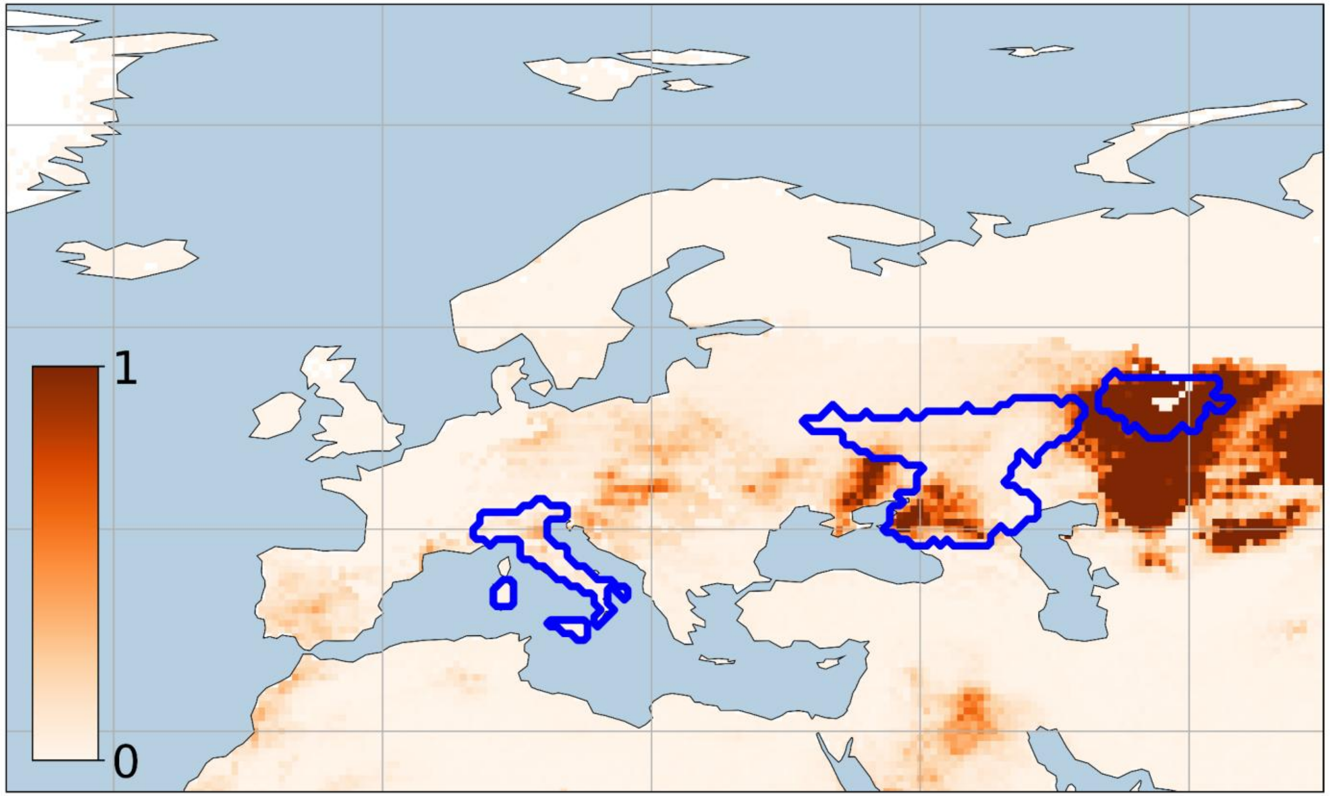
Physical Knowledge + Uncertainty

Flexible representation
Strong generalization
Statistical reliability
Physical consistency

Study 1: Extreme Drought Event Detection

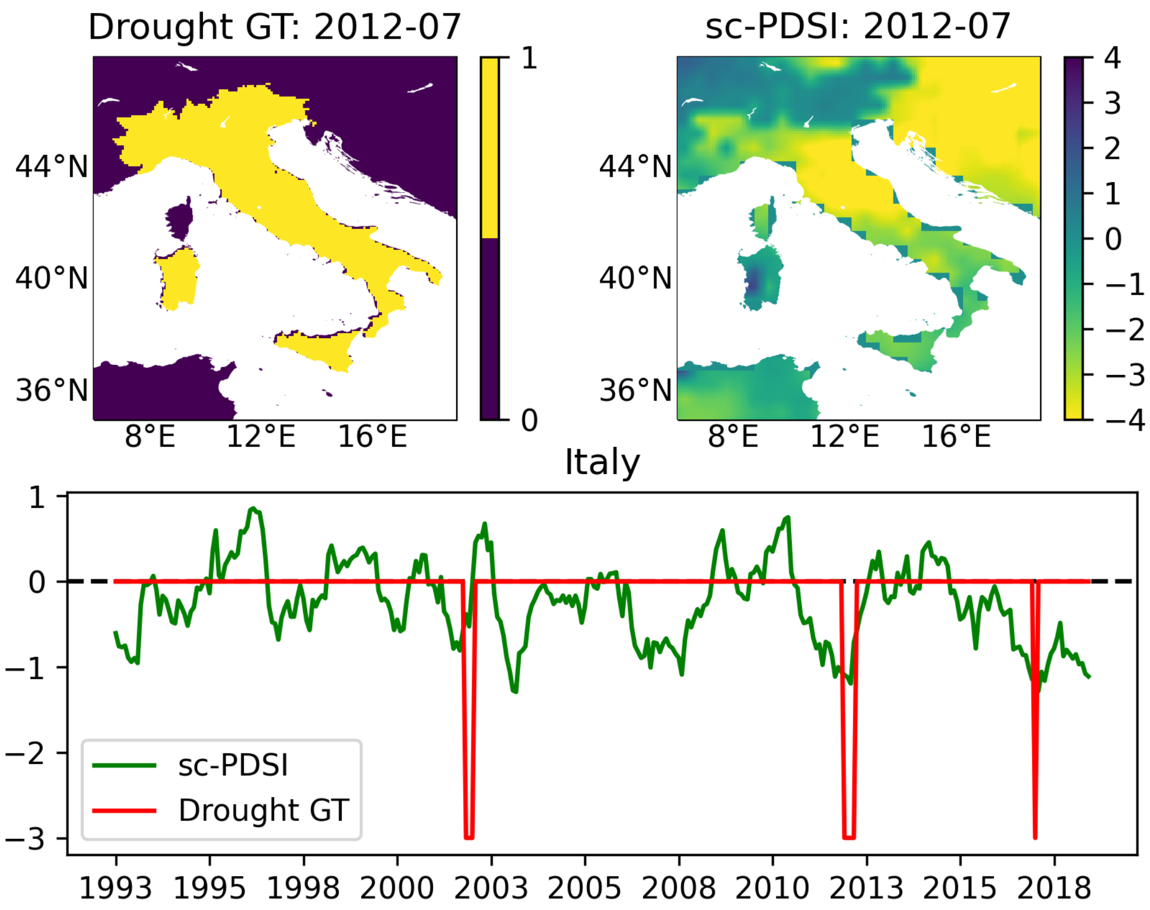
Mengxue Zhang, Miguel-Ángel Fernández-Torres, and Gustau Camps-Valls, Domain knowledge-driven variational recurrent networks for drought monitoring. *Remote Sensing of Environment*, vol. 311, pp. 114252, 2024.

1.1 Challenges – Data Imbalance



Drought events: Highly rare in space and time

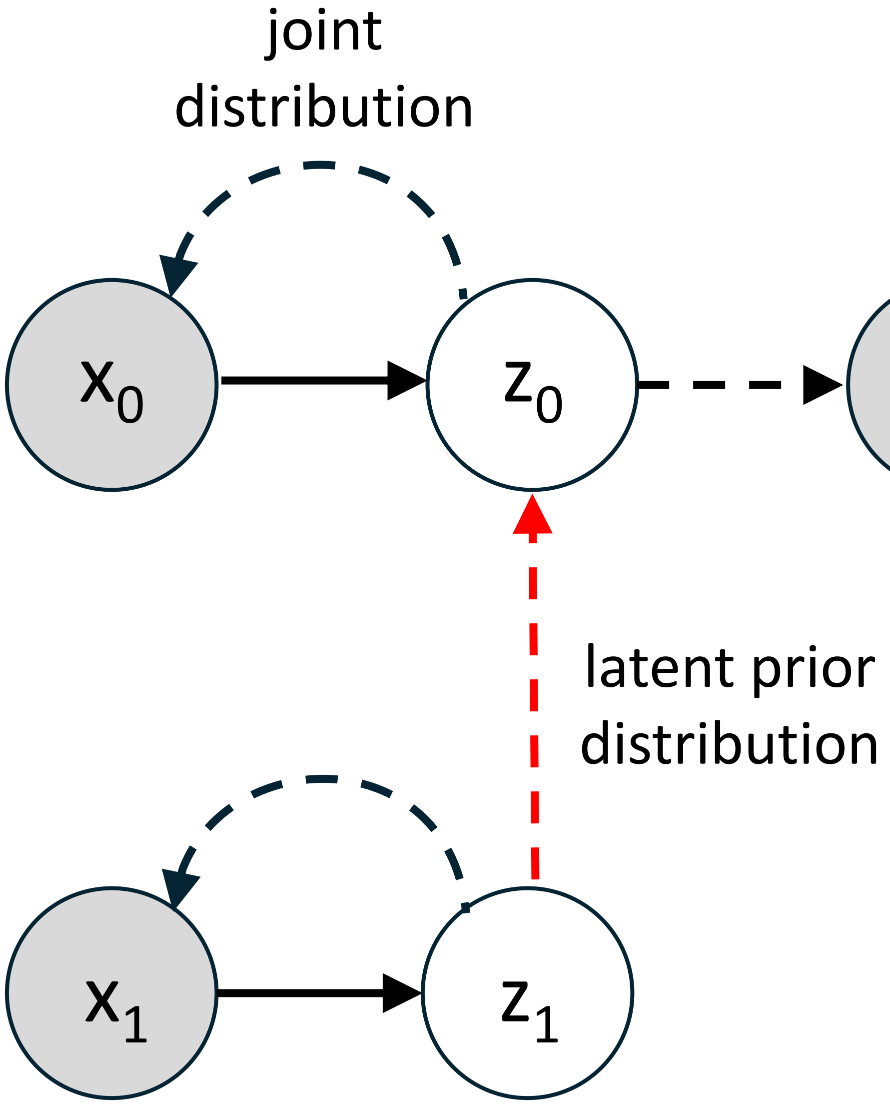
1.1 Challenges – False Alarms



sc-PDSI: Self-Calibrated Palmer Drought Severity Index

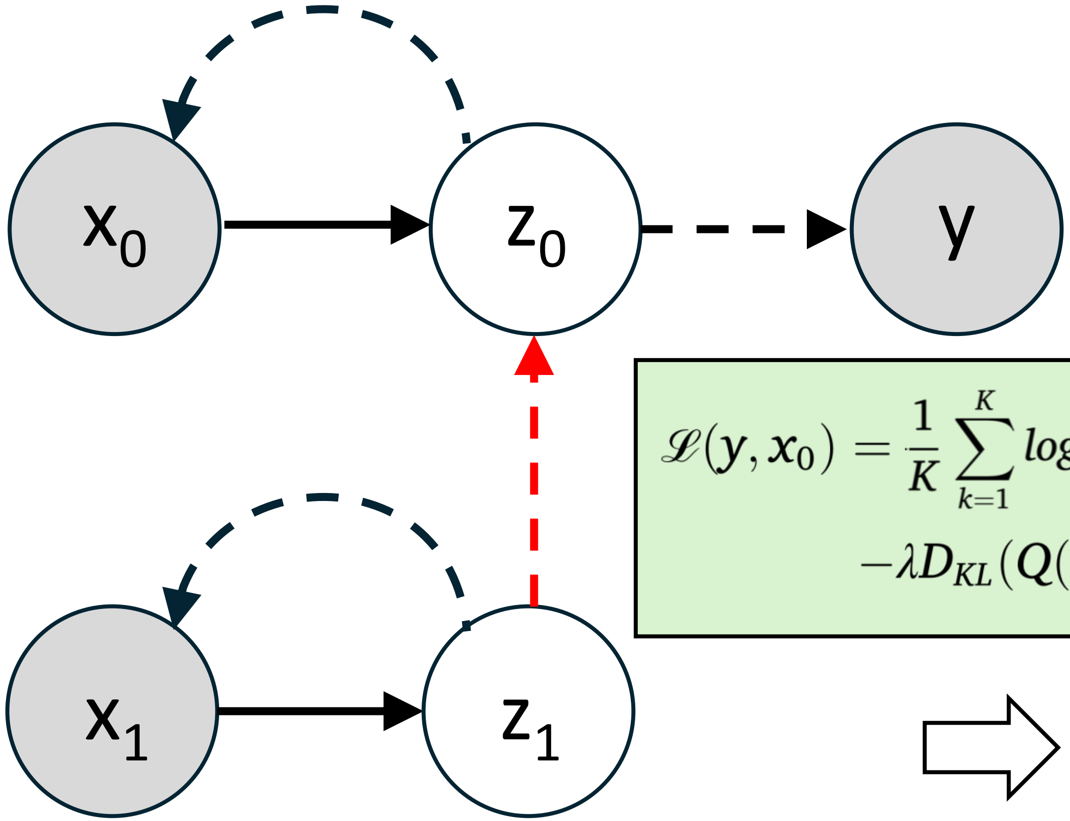
Drought indices may not distinguish drought events from internal variability

1.2 Methodology



x_0: climate variables
x_1: drought indices
z: latent variables
y: drought labels

1.3 Core Loss Function

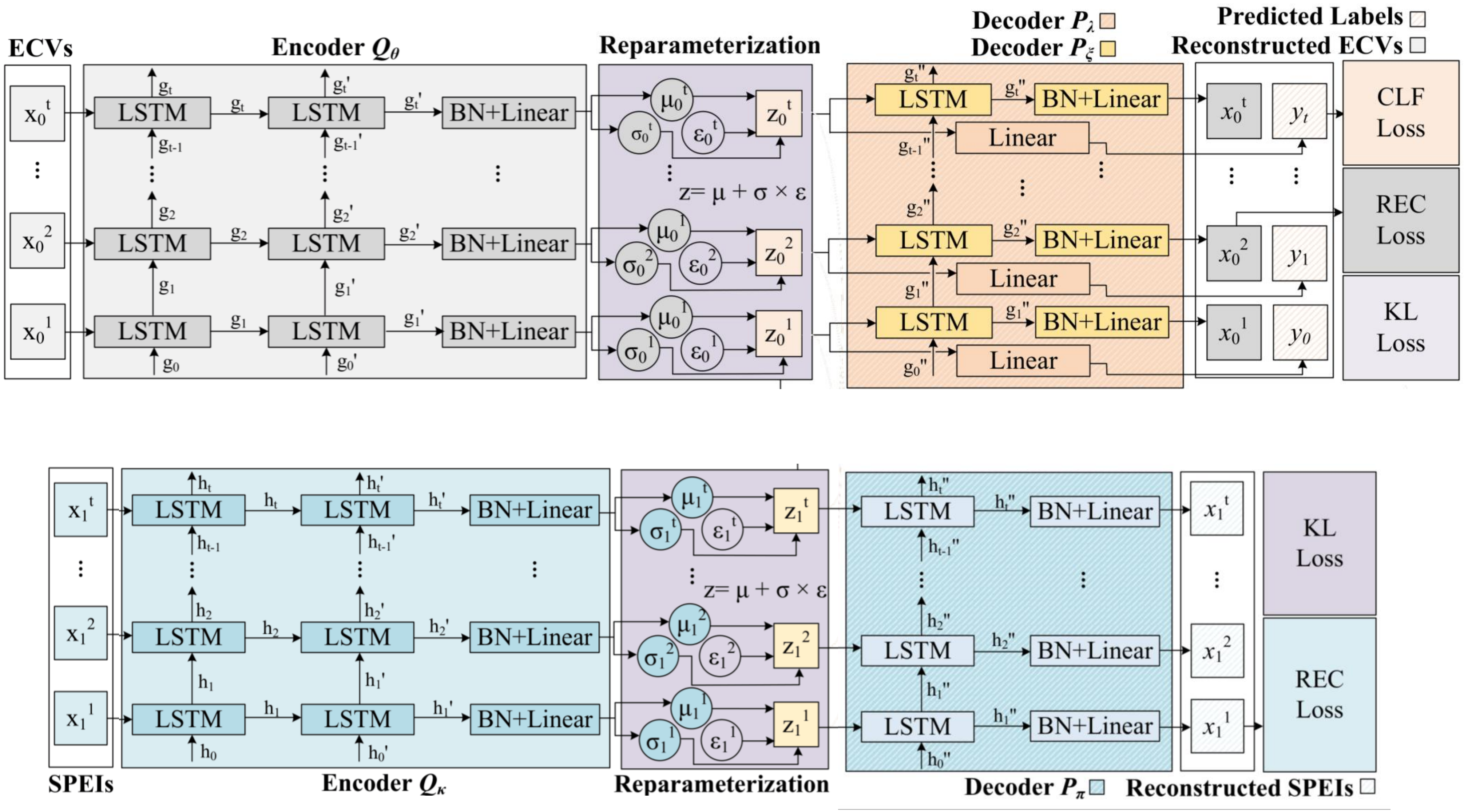


$$\mathscr{L}(y,x_0) = \frac{1}{K}\sum_{k=1}^{K} logP\left(y|x_0,z^{(k)}\right) + \frac{1}{K}\sum_{k=1}^{K} logP\left(x_0|z^{(k)}\right) - \lambda D_{KL}\left(Q(z|x_0)\|Q(z|x_1)\right)$$

$$\mathscr{L} = \mathscr{L}(y,x_0) + \mathscr{L}(x_1)$$

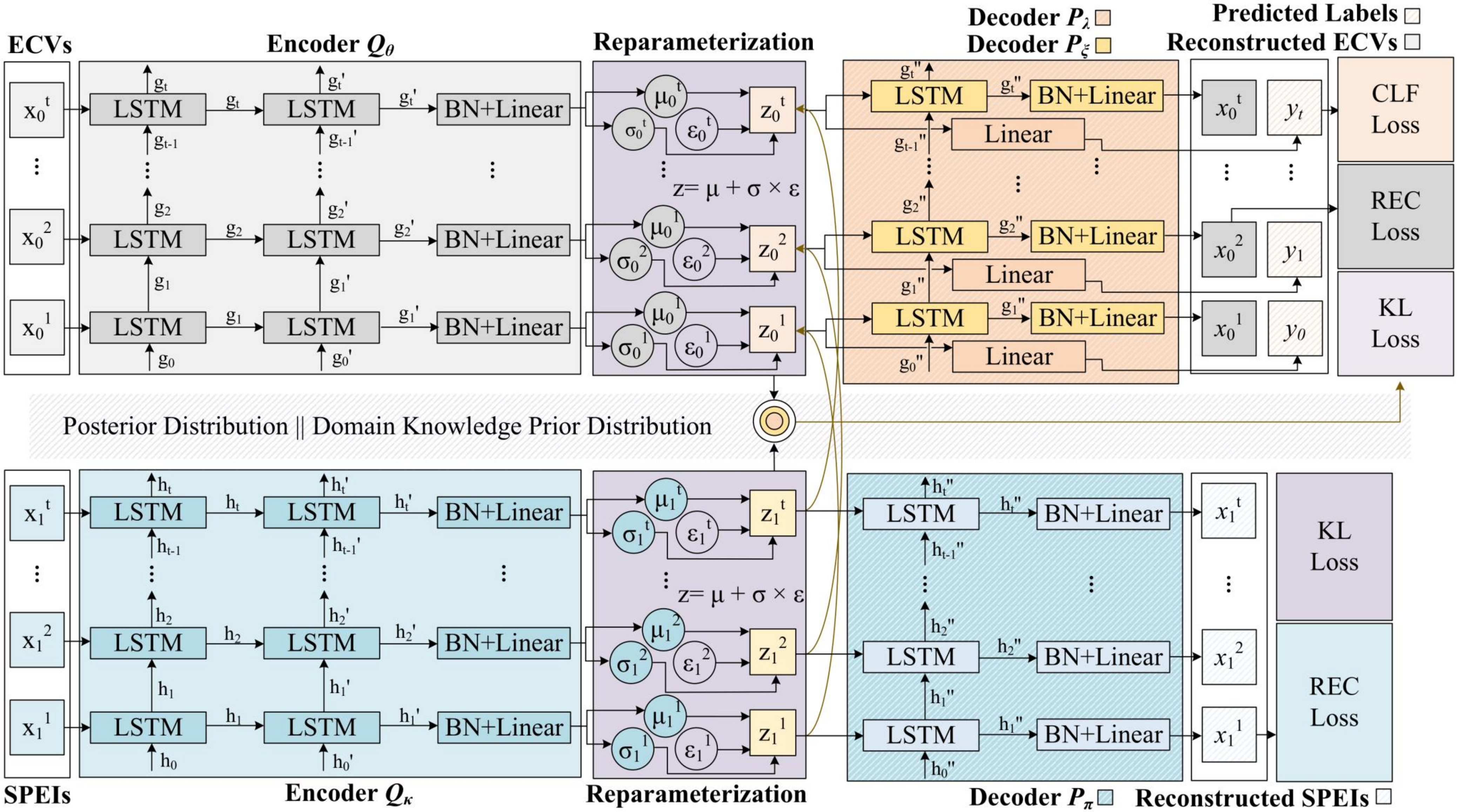
$$\mathscr{L}(x_1) = -D_{KL}\left(Q(z|x_1)\|P(z)\right) + \frac{1}{K}\sum_{k=1}^{K} logP\left(x_1|z^{(k)}\right)$$

1.4 Architecture – Domain Knowledge-Driven Variational Recurrent Networks



Note: x_0 is TP & T2M, x_1 is Multi-scalar SPEI

1.4 Architecture – Domain Knowledge-Driven Variational Recurrent Networks



1.5 Experimental Setup

(1) Data Split: Drought event subsets considered in this study.

	Training	Validation	Test
Afghanistan	2006.07–12	–	–
	2008.10–11	–	2011.01–08
Bosnia*	2003.02–05	–	–
Croatia*	2003.02–05	–	–
Hungary*	2003.02–05	–	–
Italy	–	–	2012.06–10
Lithuania	2006.08–12	–	–
Mauritania	2003.01–12	–	–
Moldova	–	–	2012.11–12
Russia	2003.01–12	2010.04–08	2012.06–08
Syria	2008.01–12	2009–2010	–
Tajikistan	2008.10–11	–	–
Period	2003–2008	2009–2010	2011–2018
Ratio (all)	0.13%	0.94%	0.15%
Ratio (happened)	9.19%	24.05%	3.80%

(2) Model baselines

MLP, RNN, LSTM
VMLP, VRNN, VLSTM, SPEI-6

(3) Performance evaluation

ROC-AUC, PR-AUC,
Macro F1, number of
parameters, running time

1.6 Quantitative Results

	Afghanistan	Italy	Moldova	Russia
SPEI-6	75.8	69.6	**97.4**	78.1
LSTM	76.3±4.3	75.7±3.4	96.5±1.2	75.6±4.1
VLSTM	74.2±3.1	68.9±7.5	95.3±2.3	80.1±1.5
DK-VRN	**79.7±0.5**	**84.3±0.6**	92.5±0.7	**89.4±0.2**

Event-based ROC-AUC (%) for different drought detection models

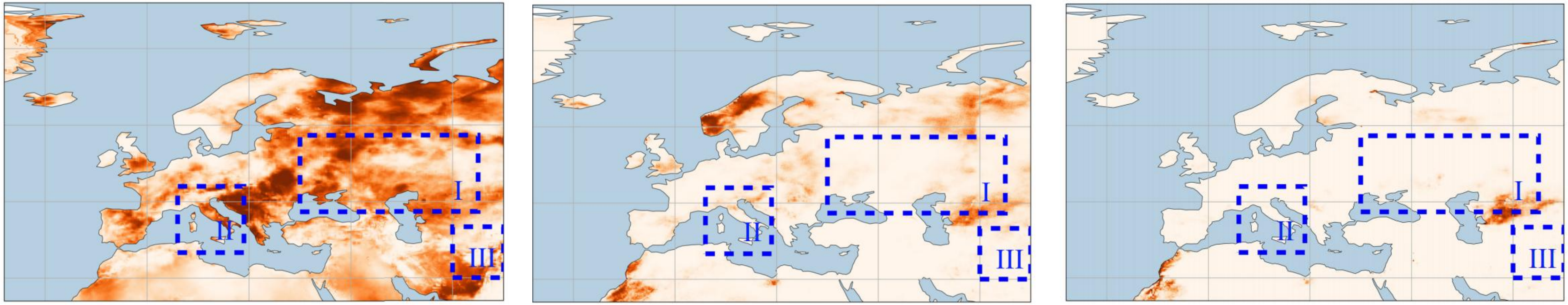
1.6 Quantitative Results

CLF-Loss	RCE-Loss	KL-Loss	ROC-AUC
✓	X	standard	74.8±3.5
✓	✓	standard	79.6±0.9
✓	X	domain knowledge	80.2±3.2
✓	✓	domain knowledge	**84.3±0.1**

Ablation study on the use of different loss terms

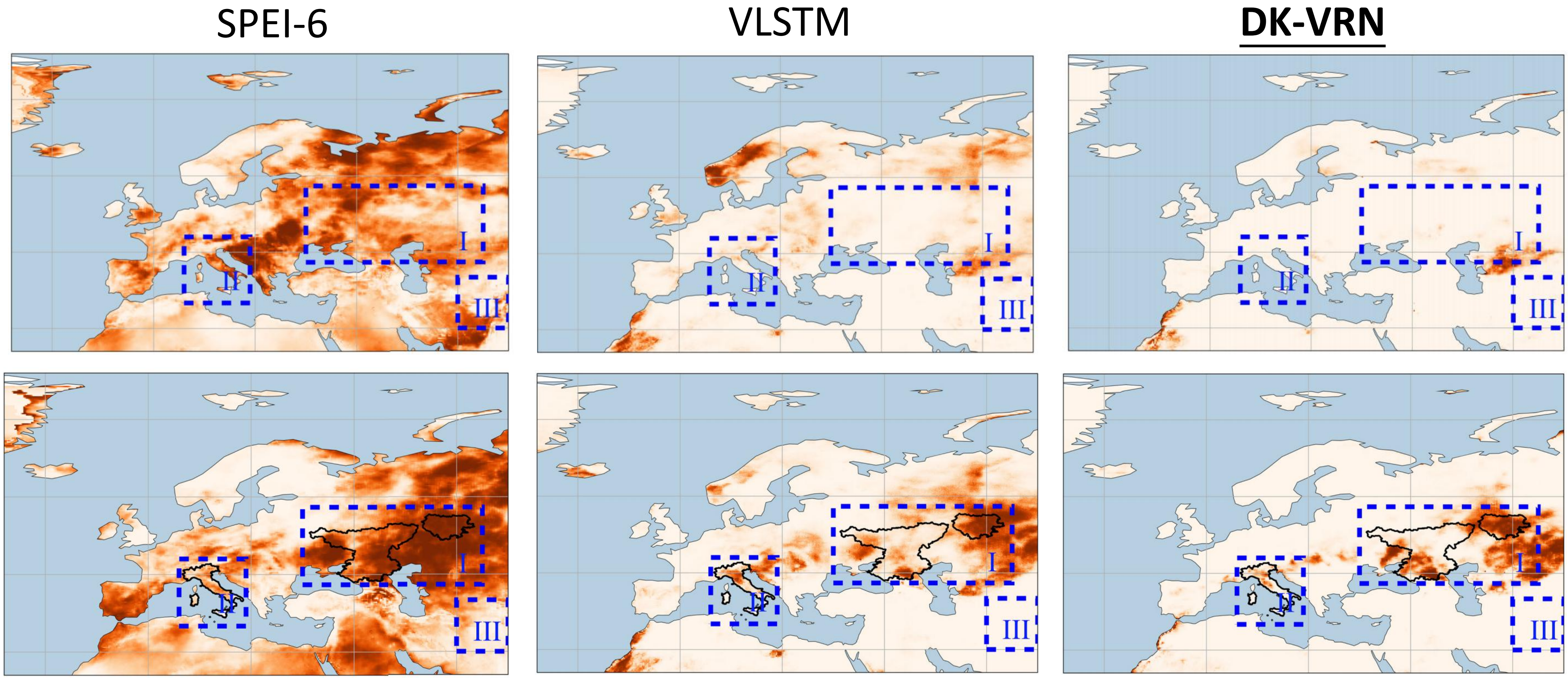
1.7 Visualizations and Qualitative Evaluation

SPEI-6 | VLSTM | **DK-VRN**



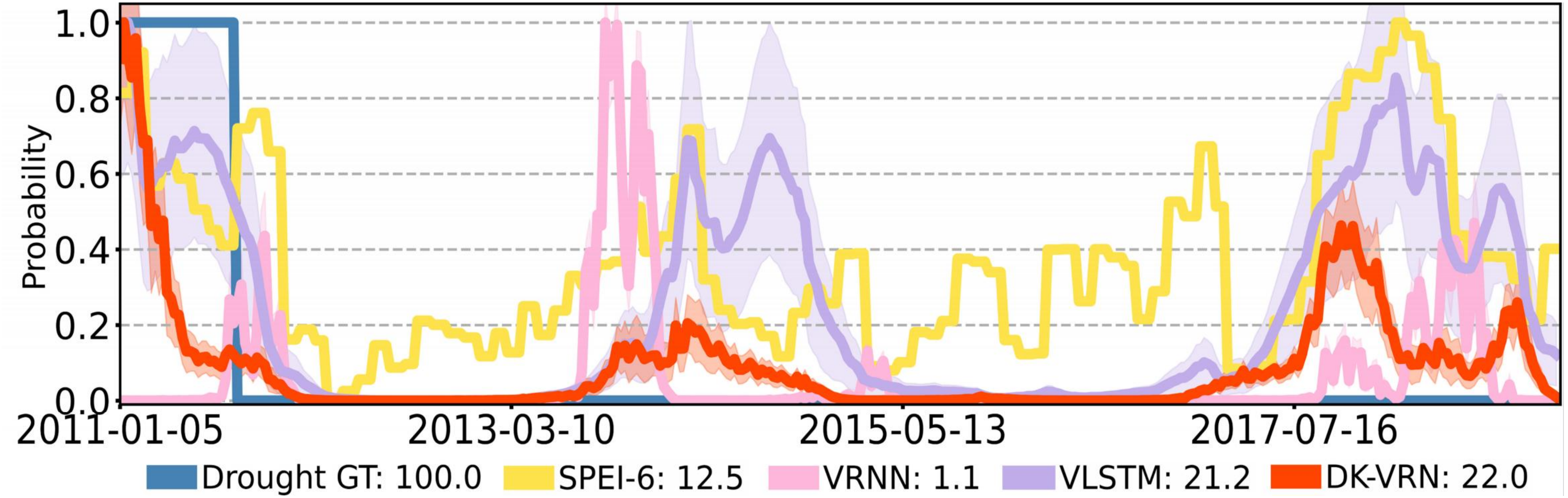
Before the drought event: Probabilistic detection maps

1.7 Visualizations and Qualitative Evaluation



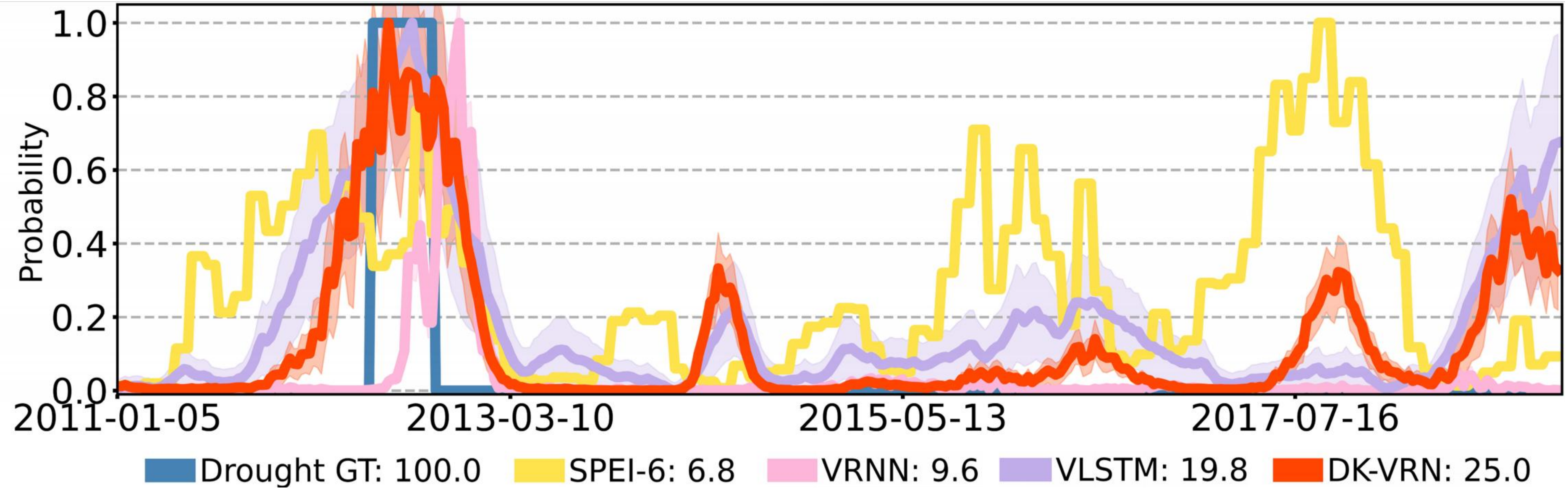
When drought occurs: Probabilistic drought detection maps

1.7 Visualizations and Qualitative Evaluation



Afghanistan: Spatially aggregated drought detection signals

1.7 Visualizations and Qualitative Evaluation



Italy: Spatially aggregated drought detection signals

1.8 Summary

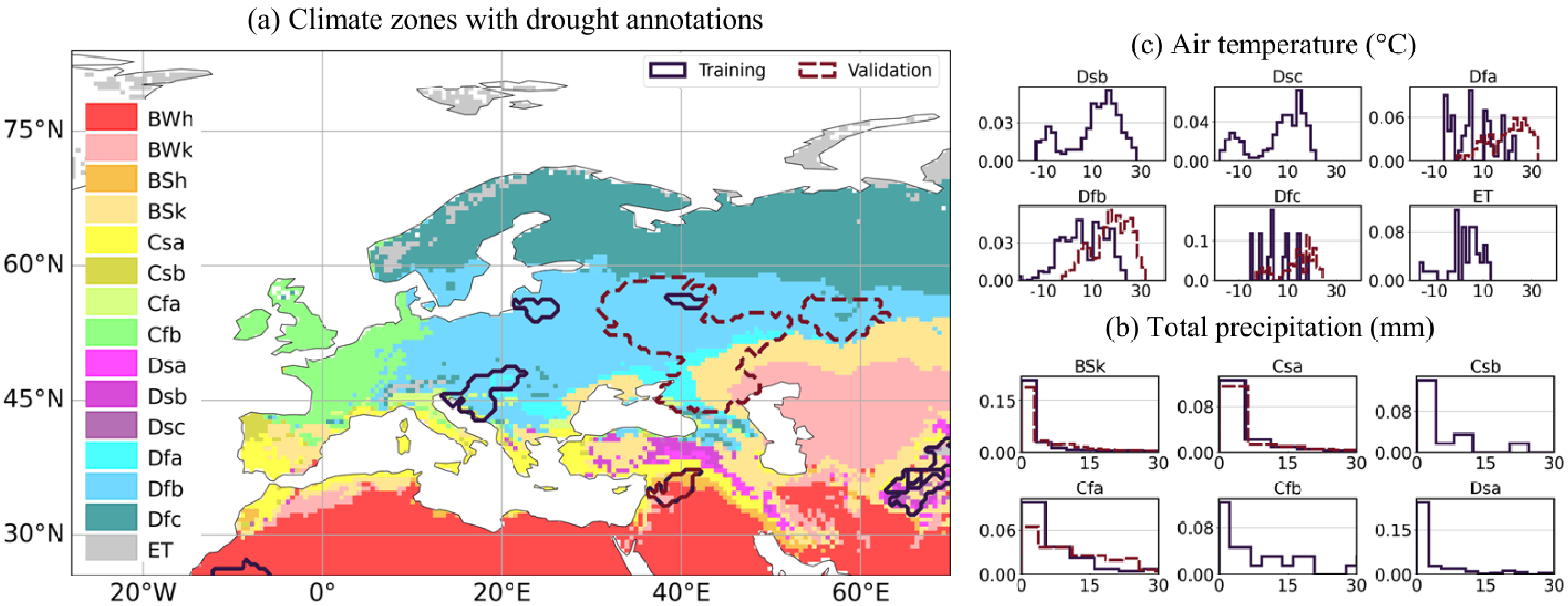
- Hybrid probabilistic model for drought detection
- Harnessing of domain knowledge as latent prior distributions
- Stronger generalization across events

Mengxue Zhang, Miguel-Ángel Fernández-Torres, and Gustau Camps-Valls, Domain knowledge-driven variational recurrent networks for drought monitoring. *Remote Sensing of Environment*, vol. 311, pp. 114252, 2024.

Study 2: Reliable Probabilistic Drought Prediction

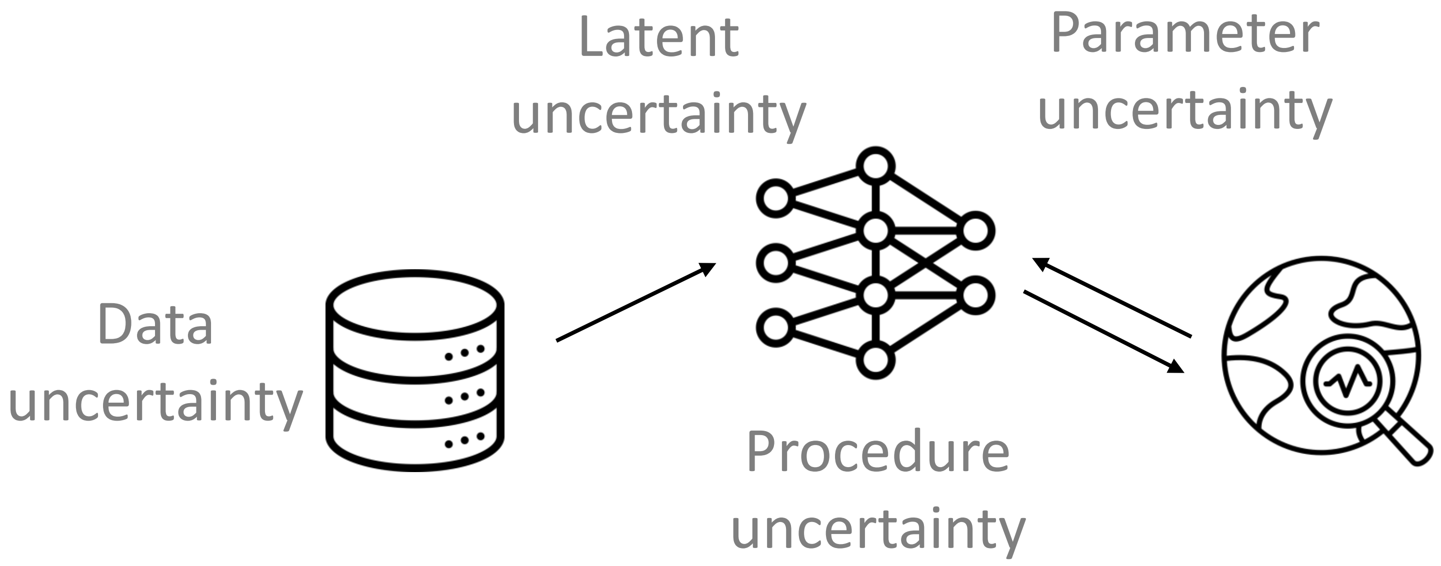
Mengxue Zhang, Miguel-Ángel Fernández-Torres, Kai-Hendrik Cohrs, and Gustau Camps-Valls, Calibration and uncertainty quantification for deep learning-based drought detection. *International Journal of Applied Earth Observation and Geoinformation*, vol. 140, pp. 104563, 2025

2.1 Challenges – Distributional Shifts



Distributional shifts often cause overconfident model predictions

2.1 Challenges – Uncertainty Quantification



Comprehensive **uncertainty quantification** is important for model robustness

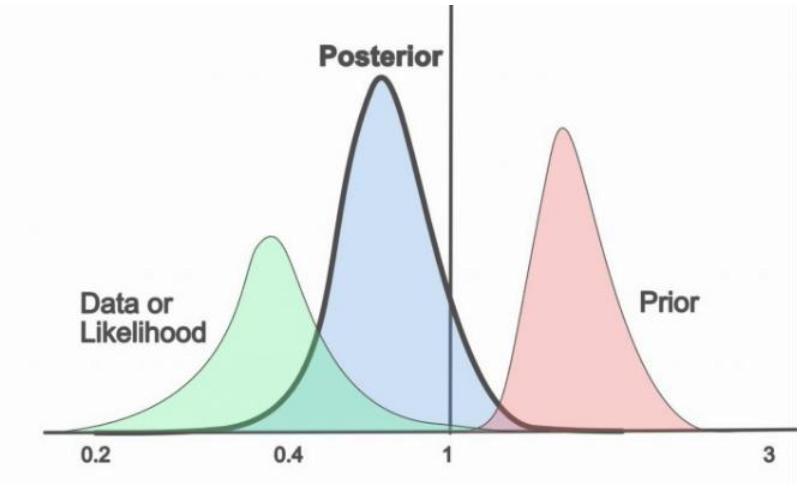
2.2 Two-Step Calibration

Observations:

Distributional shifts include label-prior and likelihood shifts

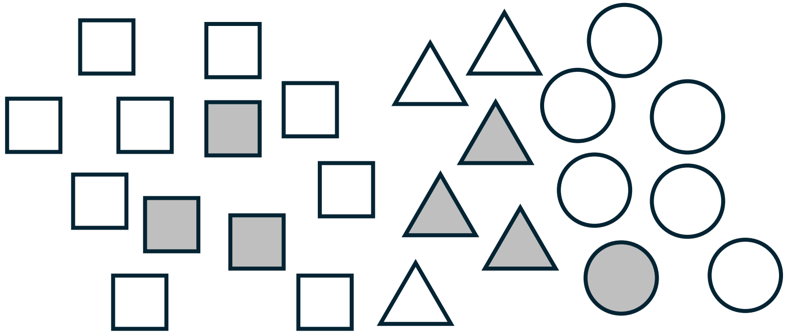
(1) Label-prior shift: different drought prior distribution of climate zones

(2) Likelihood shift: between training and validation subsets

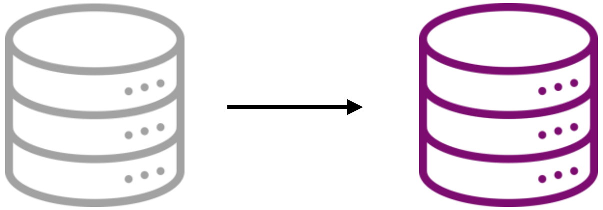


Bayes' theorem

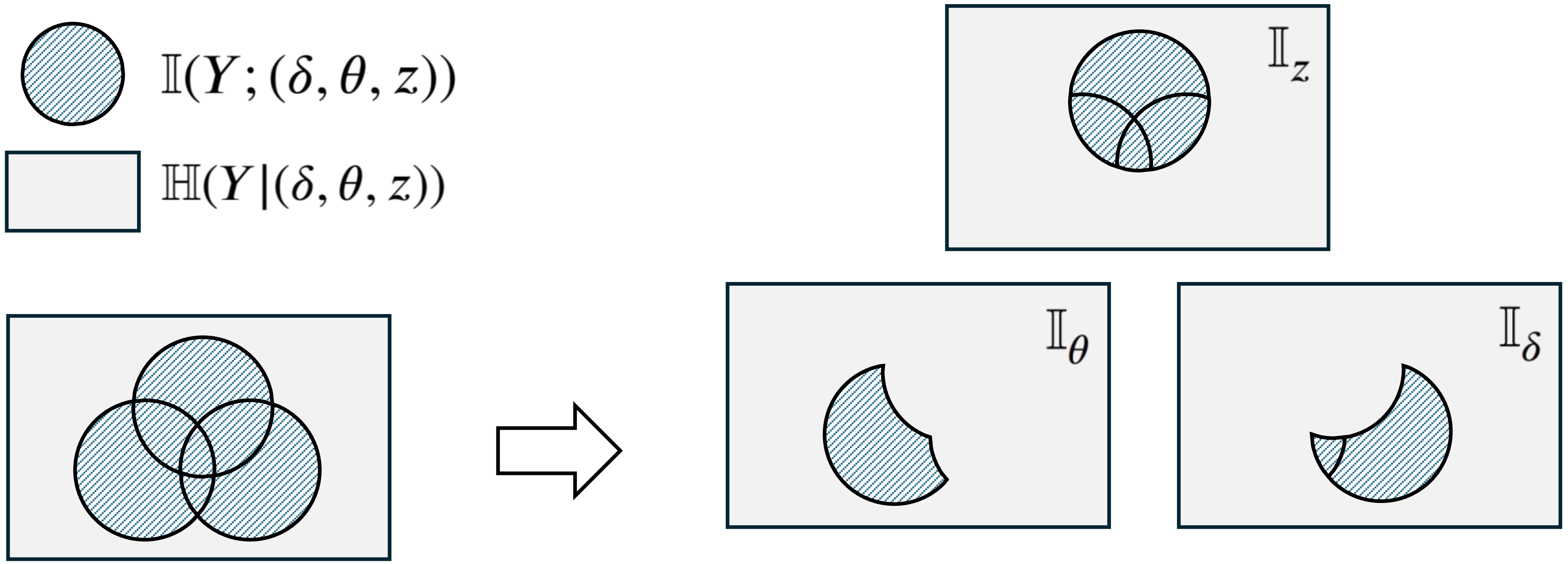
Calibrated functions



Bias corrections



2.3 Entropy-based Uncertainty Quantification



Total uncertainty
(model I, data H)

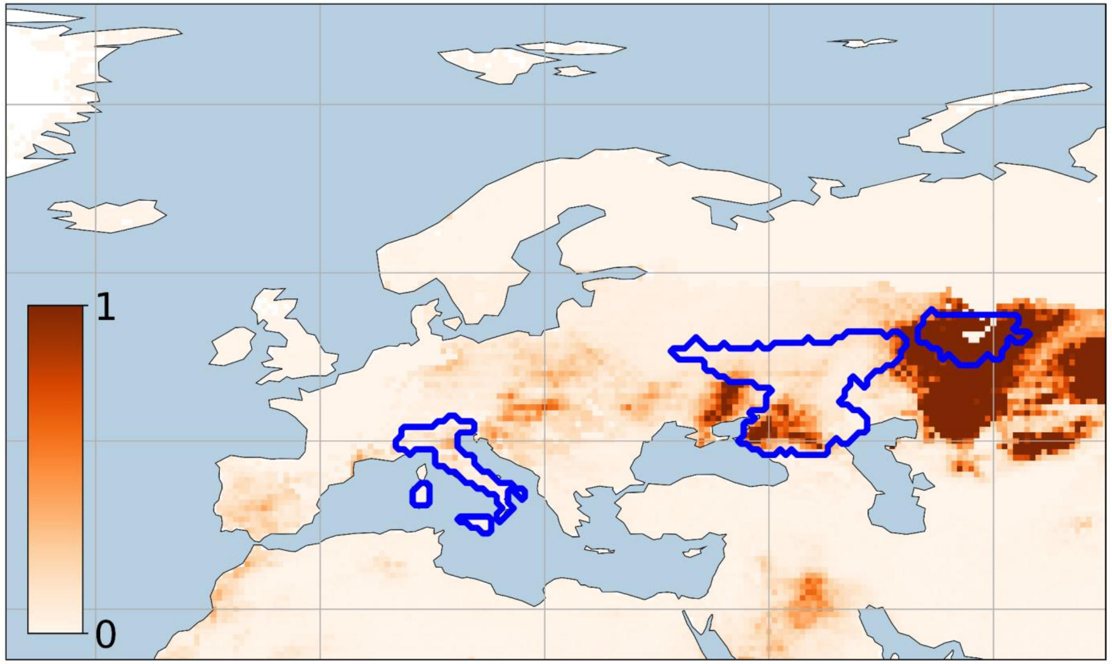
Model uncertainty
(latent z, parameter θ, procedure δ)

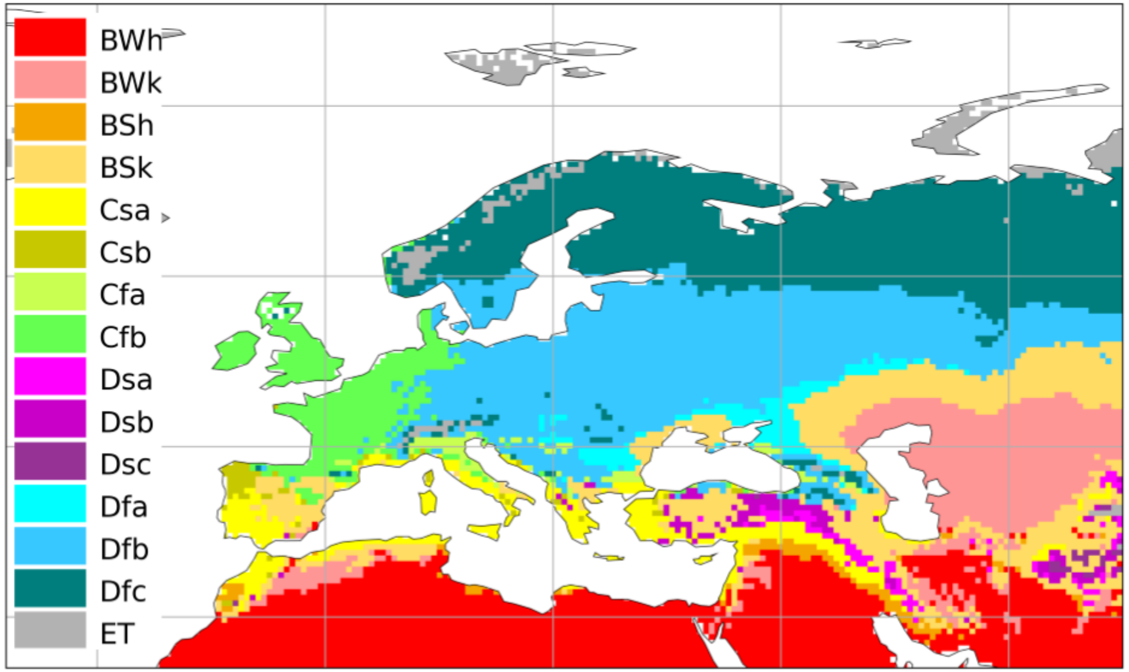
Different sampling

Monte Carlo dropout

Random seeds

2.4 Experimental Setup





(1) Dataset

ERA5 (features),
EM-DAT (label),
Climate zones

(2) Calibration baselines

Temperature scaling,
Platt scaling,
Polynomial scaling

(3) DL-based predictions

VLSTM, DK-VRN (Study 1)

(4) Evaluation metrics

ROC-AUC, PR-AUC,
Macro F1, ECE, RMSCE

2.5 Calibration Errors

	Uncalibrated	Temp	Platt	Ploy	Proposed
ECE(↓)	4.06	3.64	0.40	0.35	**0.31**
RMSCE(↓)	3.40	2.48	0.32	0.28	**0.23**
Training (s)	-	17.8	30.41	3.17	72.0
Inference (s)	-	0.07	0.07	0.03	0.07

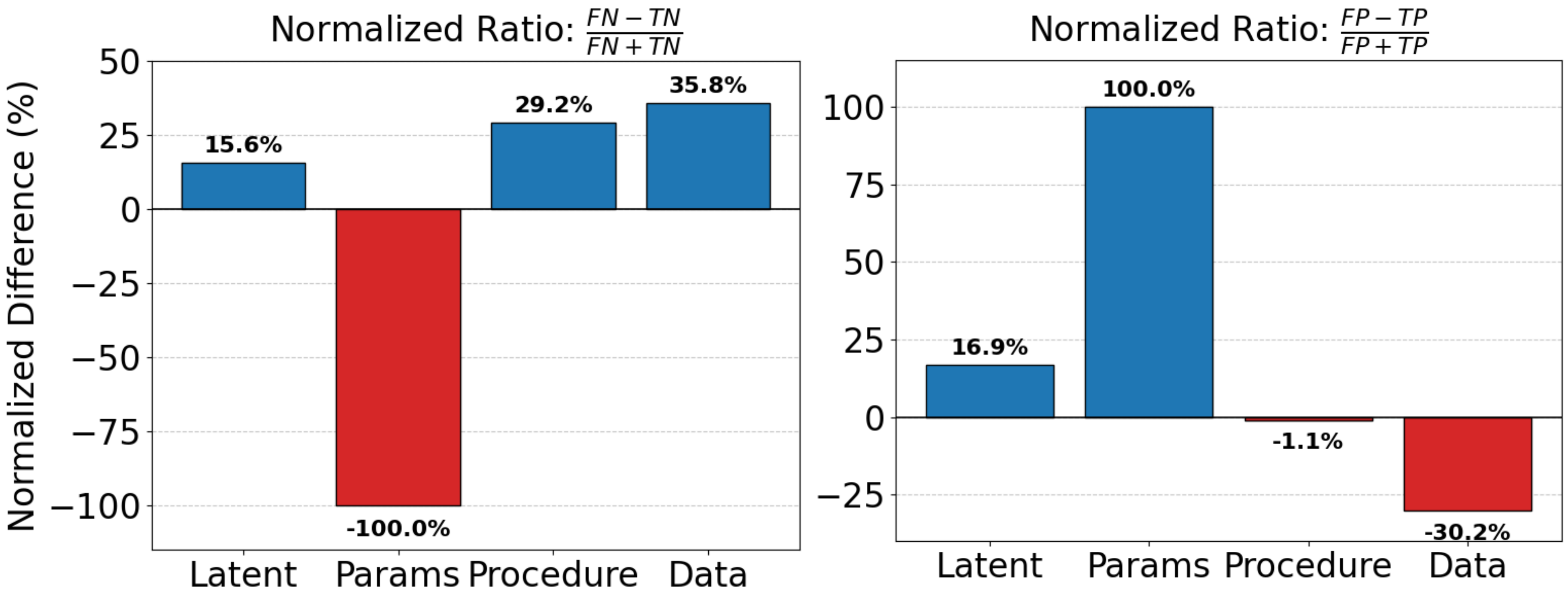
Overall calibration errors (%) for different parametric calibration algorithms.

2.6 Uncertainty Analysis – Total Uncertainty

	Corrected samples			Erroneous samples		
	TP + TN	TP	TN	FP + FN	FP	FN
Total	**1.74**	31.33	1.74	**17.14**	18.24	3.77

Quantification of the different sources of uncertainty (%) for DK-VRN

2.6 Uncertainty Analysis – Normalized Difference of Uncertainty



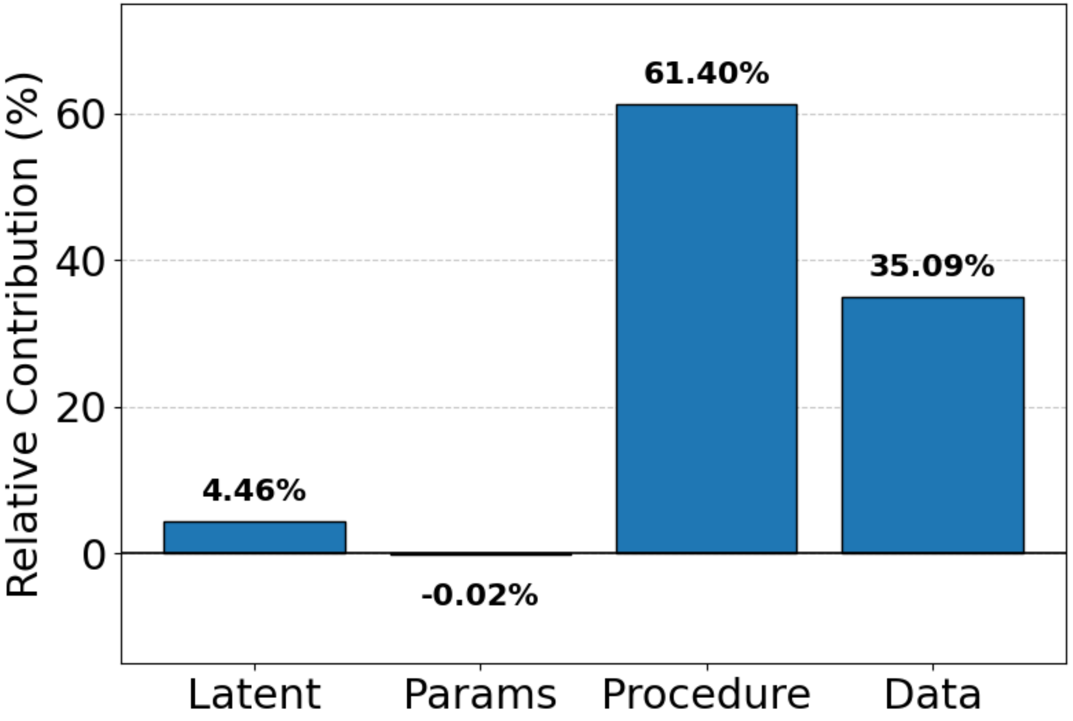
Quantification of the different sources of uncertainty (%) for DK-VRN

2.6 Uncertainty Analysis – Uncertainty between Models

	Training		Test	
	VLSTM	**DK-VRN**	VLSTM	**DK-VRN**
Total	3.09	**3.02**	3.83	**3.26**

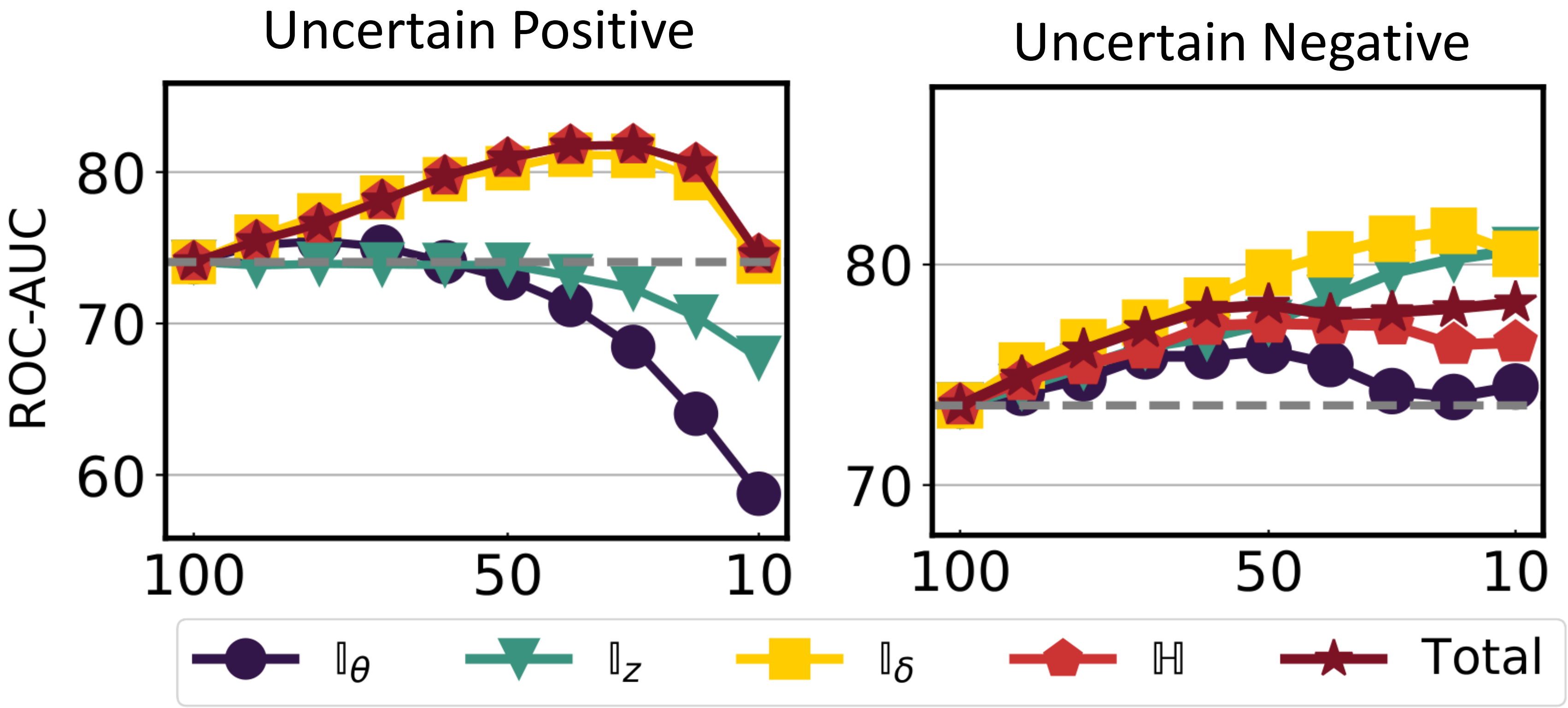
Quantified uncertainty (%) of different models
(VLSTM and DK-VRN, in Study 1)

2.6 Uncertainty Analysis – Relative Contribution



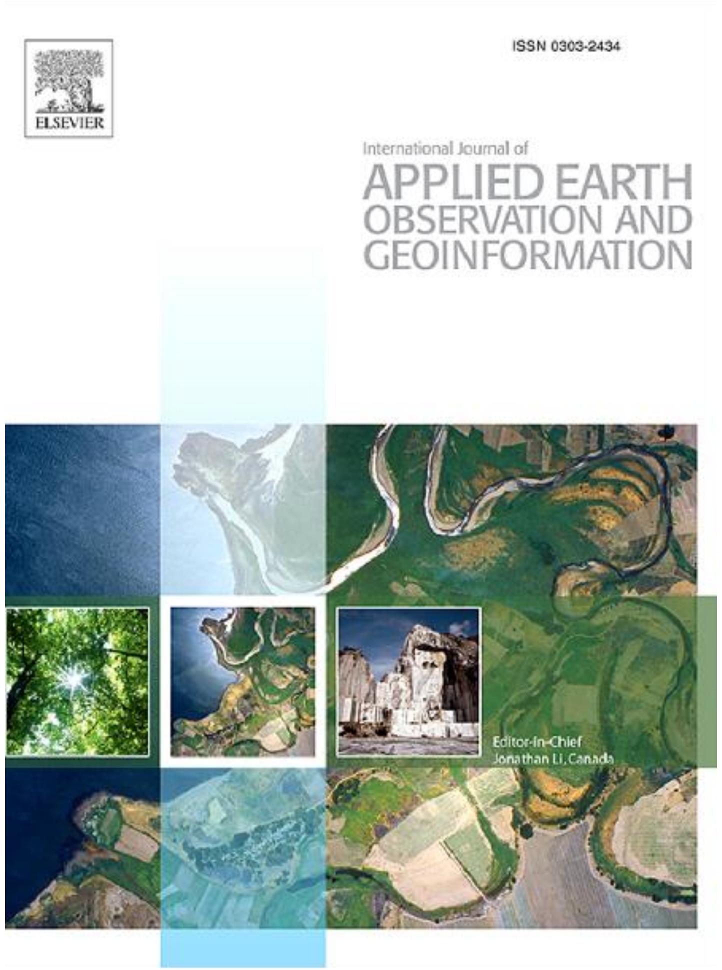
Quantified uncertainty (%) of different models
(VLSTM and DK-VRN, in Study 1)

2.6 Uncertainty Analysis – Decision-Making



Accuracy–rejection curves (%), given uncertain positive and negative samples. The x-axis denotes the keeping rate from 100% to 10%.

2.7 Summary



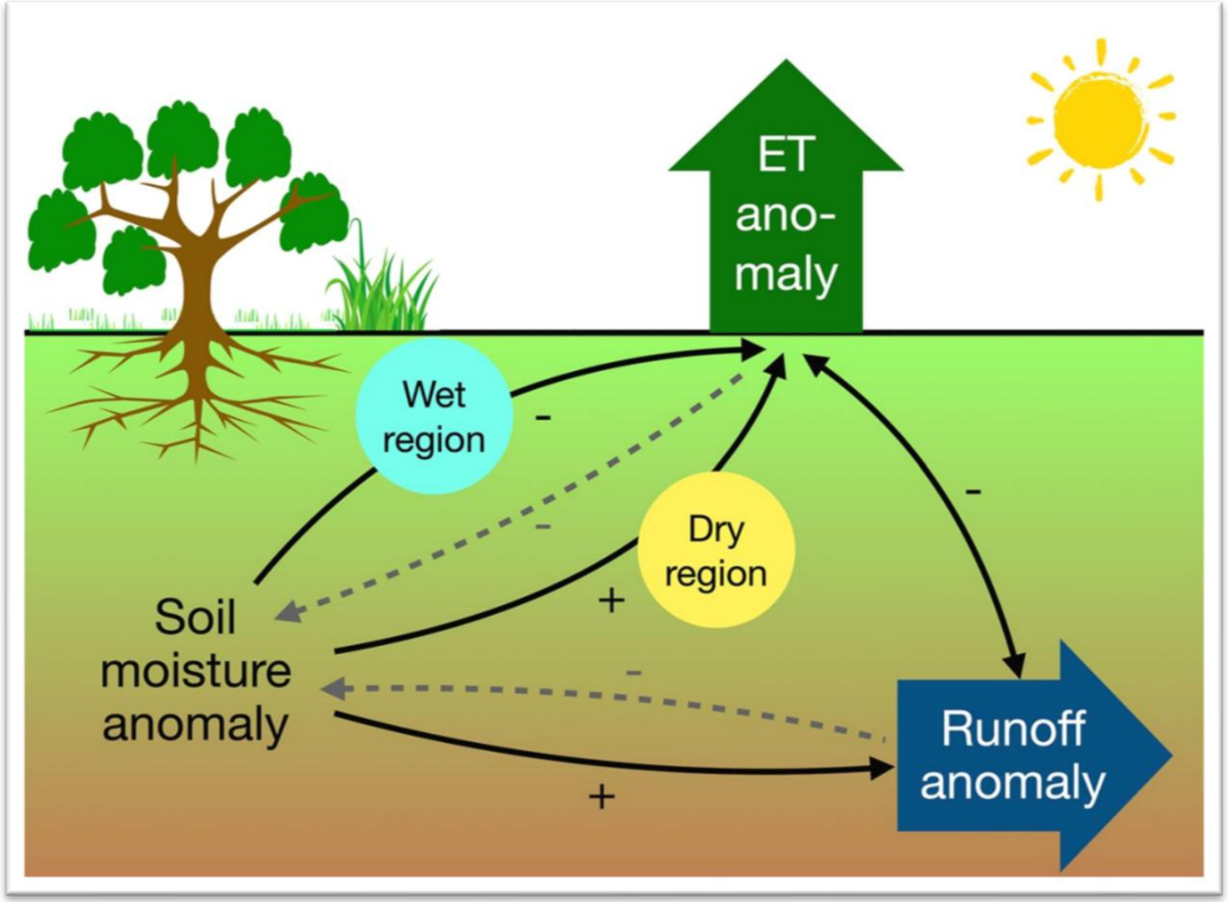
- Calibration algorithm addresses distributional shifts
- Entropy-based framework decomposes the total uncertainty
- Improve the reliability of predictions

Mengxue Zhang, Miguel-Ángel Fernández-Torres, Kai-Hendrik Cohrs, and Gustau Camps-Valls, Calibration and uncertainty quantification for deep learning-based drought detection. *International Journal of Applied Earth Observation and Geoinformation*, vol. 140, pp. 104563, 2025

Study 3: Long-term Soil Moisture Drought Forecasts

Mengxue Zhang, Miguel-Ángel Fernández-Torres, and Gustau Camps-Valls, Long-Term Soil Moisture Drought Forecasts Using a Robust Deep Learning Framework, *Under Review, 2025*

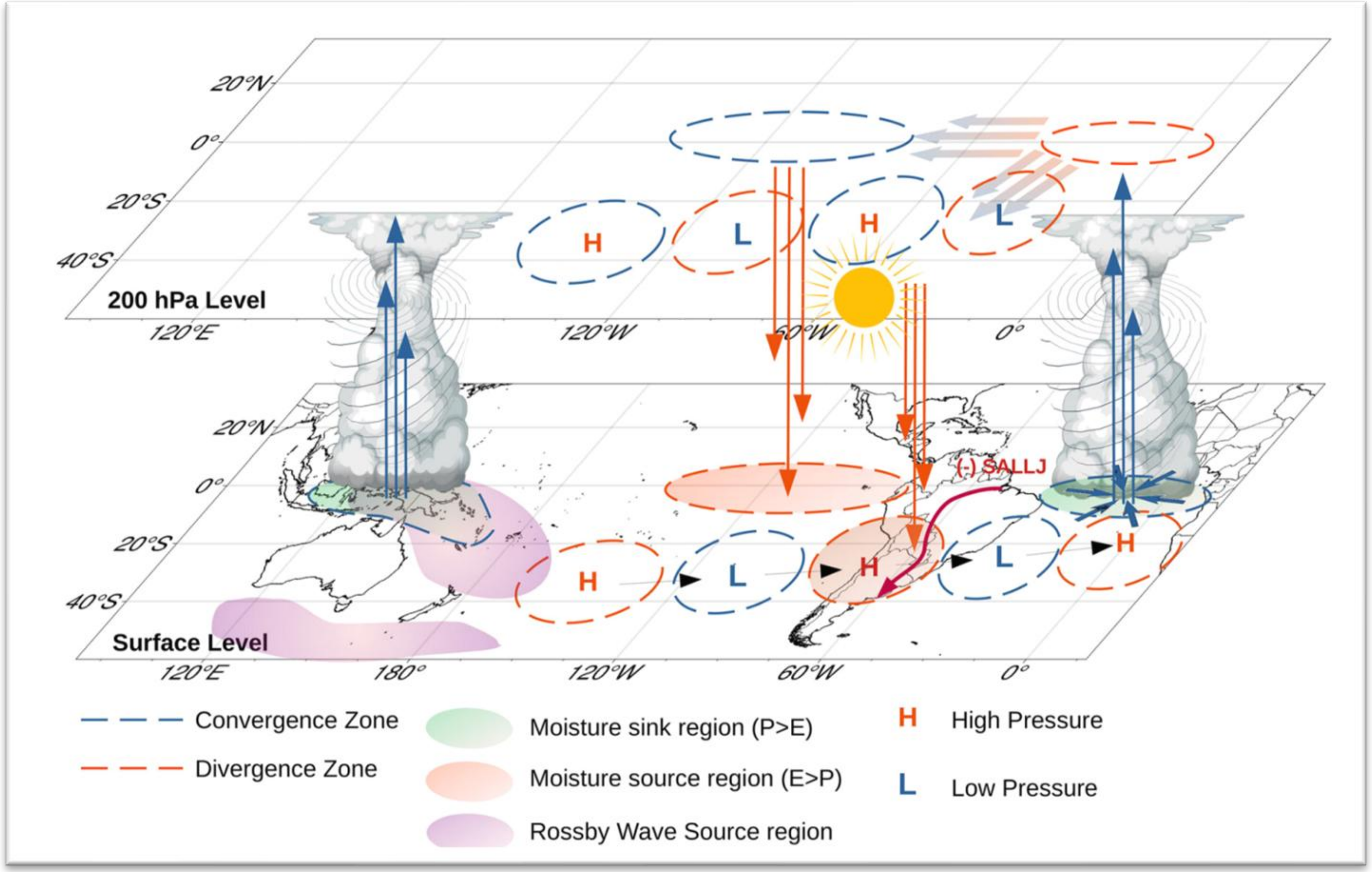
3.1 Challenges – Nonlinear Dynamics



Data source:
doi.org/10.1029/2022EF003441

At local scale, soil moisture evolves through complicated interactions

3.1 Challenges – Internal Variability



Data source:
doi.org/10.1038/s41612-023-00510-3

At large-scale, soil moisture dynamics are modulated by internal climate variability

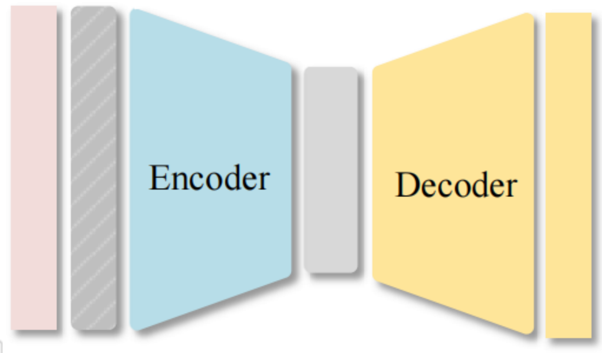
3.2 The Framework

Concepts:

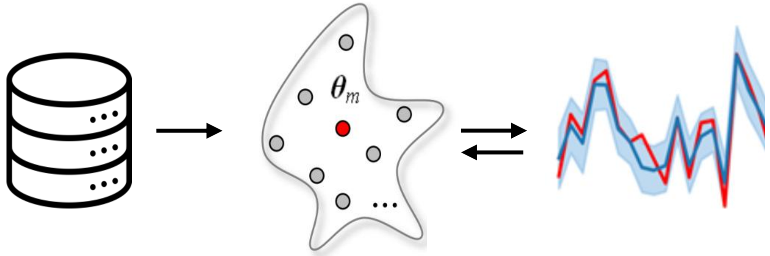
(1) Autoregressive components for nonlinear processes

(2) Climate variability can be modelled with more data

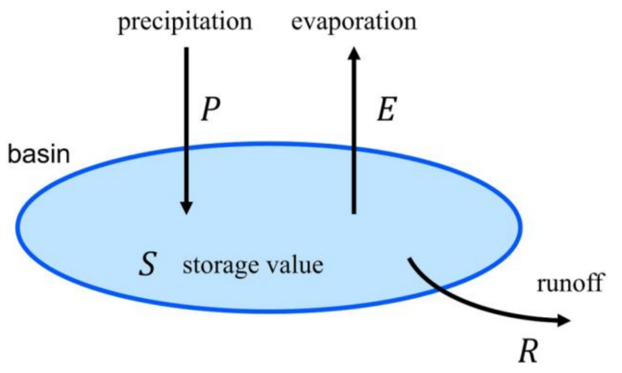
(3) Incorporating the moisture budget increases the physical consistency



MaskTRN Architecture

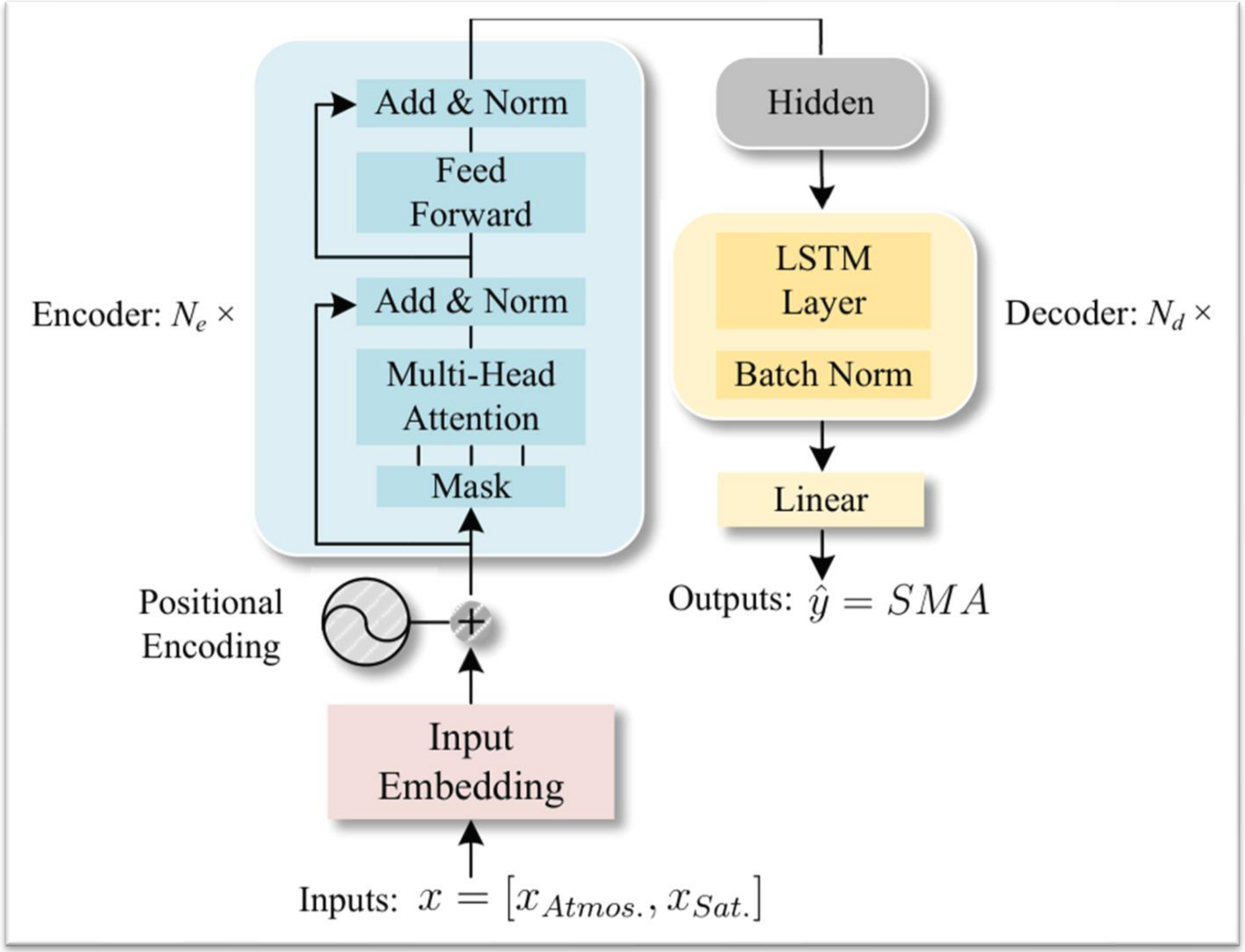


Data-aware Uncertainty



Physics-aware Regularization

3.2 MaskTRN Architecture



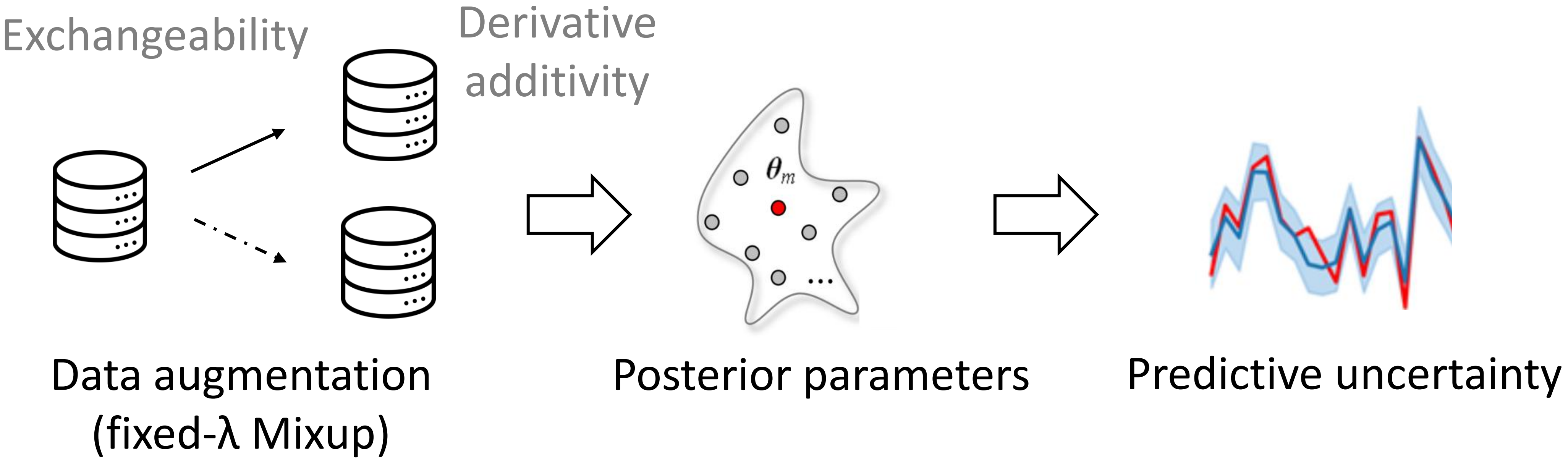
The diagram of the proposed MaskTRN

3.3 Data-aware Uncertainty (DaU)

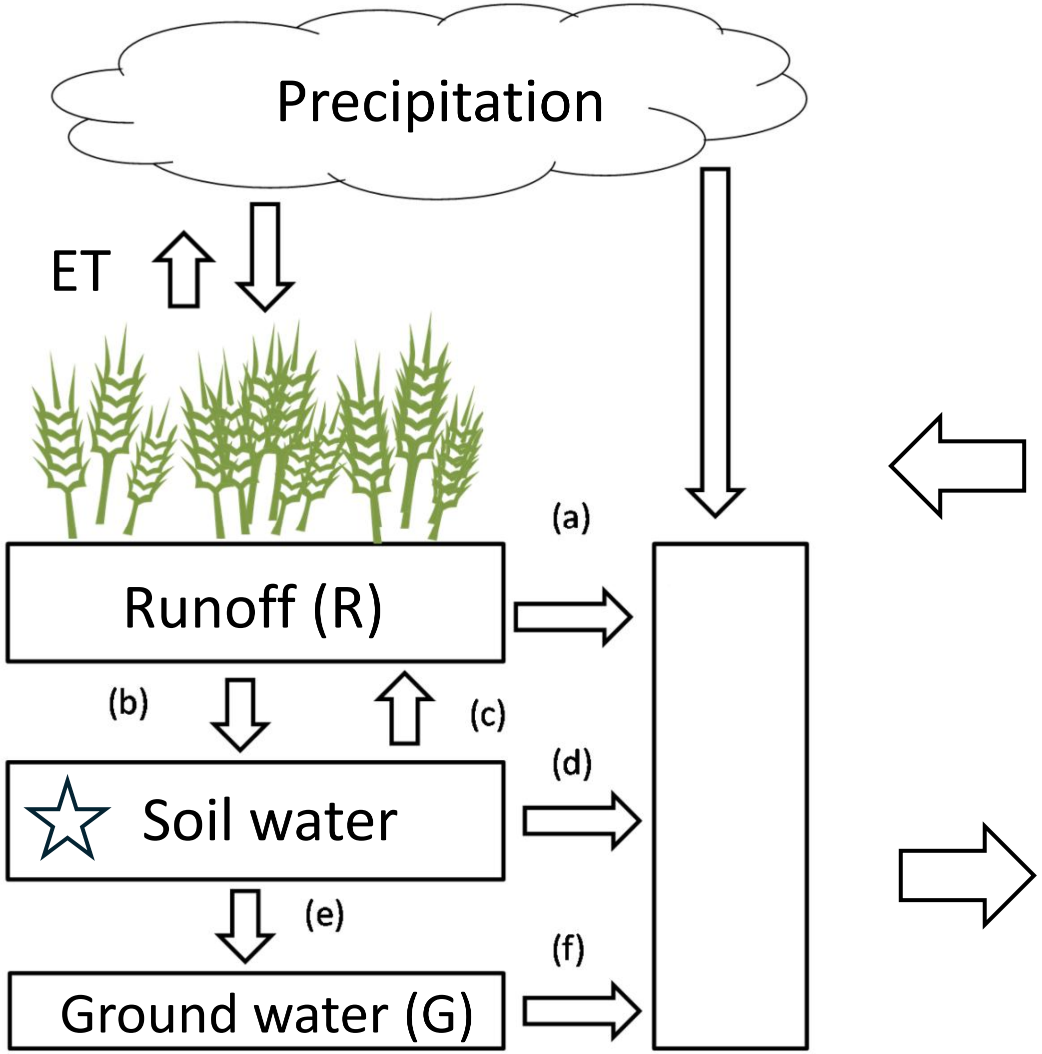
Martingale posterior:

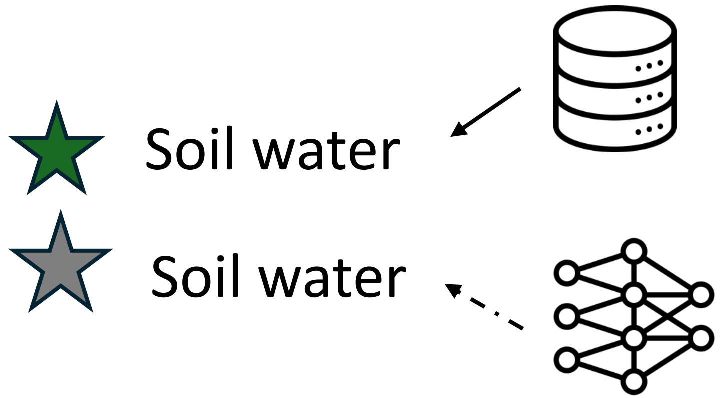
(1) Generates predictive distribution by estimating future data distribution, under the infinite exchangeability assumptions

(2) For DL models, the estimation can be realized with the well-defined data augmentation and deep ensembles



3.4 Physics-aware Regularization (PaR)





$$V(t) = [E(t), R(t), G(t)]$$
$$\mathcal{L}_\Phi = \|V(t) - \tilde{V}(t)\|^2$$

3.5 Experimental Setup

(1) Data set:

Atmospheric: TP, T2M, SSRD, WS, etc.
Land-surface: LAI, Albedo, LC, etc.
Target Variables: **LSM** SMA, **Satellite** SMA, 0.25°
Time Span: **1981-2008, 2009-2018**, Monthly

(2) Evaluation metrics

Generalization Error: RMSE, MAE, and R2
Calibration Error: RMSCE, MACE, and MA
Consistency Error: Pearson Correlation Coefficient, Nash-Sutcliffe Efficiency

(3) Model baselines

Random Forest (RF), Extreme Gradient Boosting (XGBoost), and DL models:

	RNN	LSTM	GRU	Transformer	MaskTRN (ours)
Layer 0	RNN_Encoding, 128	LSTM_Encoding, 128	GRU_Encoding, 128	Input_Embedding, 128	Input_Embedding, 128
Layer 1	RNN_Encoding, 128	LSTM_Encoding, 128	GRU_Encoding, 128	Positional_Encoding, 128	Positional_Encoding, 128
Layer 2	Linear, 128	Linear, 128	Linear, 128	Transformer_Encoding, 128	Transformer_Encoding, 128
Layer 3	Linear, 128	Linear, 128	Linear, 128	Transformer_Encoding, 128	Transformer_Encoding, 128
Layer 4	**Linear, 1**	**Linear, 1**	**Linear, 1**	Transformer_Encoding, 128	Transformer_Encoding, 128
Layer 5	-	-	-	Transformer_Encoding, 128	Transformer_Encoding, 128
Layer 6	-	-	-	Transformer_Encoding, 128	Transformer_Encoding, 128
Layer 7	-	-	-	**Linear, 1**	LSTM_Decoding, 128
Layer 8	-	-	-	-	LSTM_Decoding, 128
Layer 9	-	-	-	-	**Linear, 2**

3.6 Quantitative Results

	LSM-based			Satellite-derived		
	R^2(↑)	MA(↓)	ρ(↑)	R^2(↑)	MA(↓)	ρ(↑)
RF	32.78	48.35	55.98	33.61	48.01	60.10
LSTM	64.67	27.51	81.15	53.13	33.71	74.13
Transformer	65.89	26.94	79.90	54.66	30.02	73.92
Proposed	**72.21**	**8.71**	**83.67**	**59.38**	**4.30**	**76.44**

Note: Proposed (MaskTRN + DaU + PaR)

Results (%) on SMA predictions given different models.

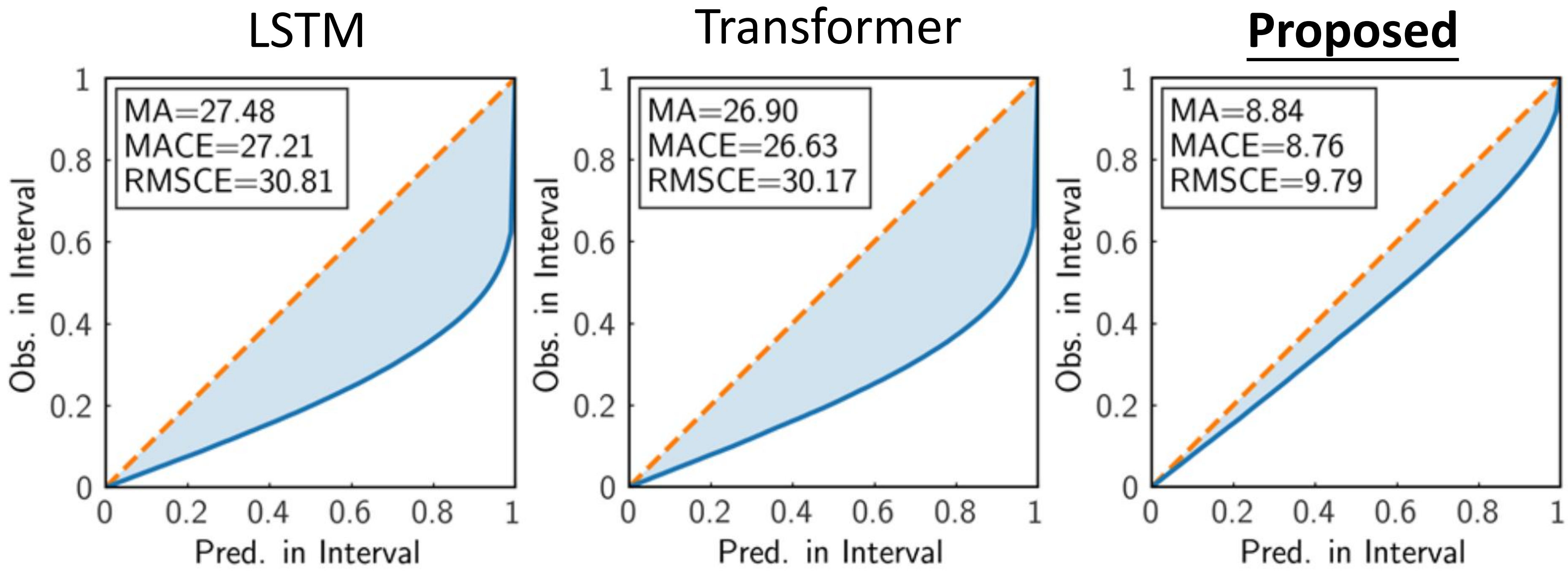
3.6 Quantitative Results

	$R^2(\uparrow)$	MA(↓)	ρ(↑)
Multi-loss	69.65	32.45	81.20
Mixup	69.29	31.73	82.06
DE	71.88	9.29	83.59
Mixup MP	71.08	9.80	83.41
Proposed	**72.21**	**8.71**	**83.67**

Note: Proposed (MaskTRN + DaU + PaR), DE (Deep Ensemble)

Results (%) on SMA predictions given different strategies for model robustness.

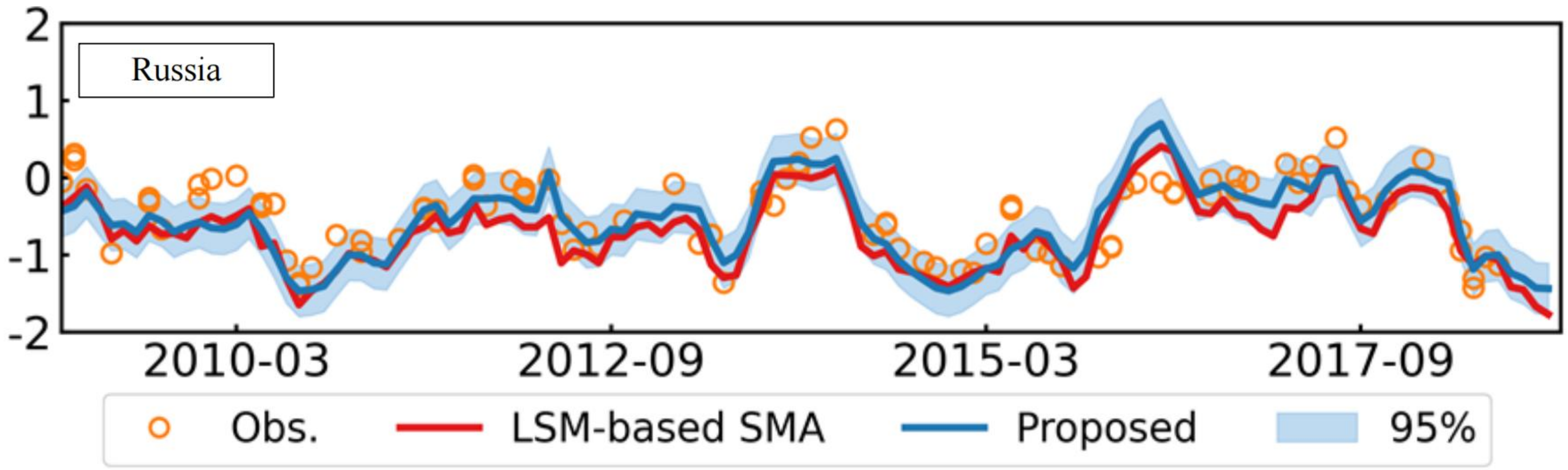
3.7 Qualitative Results



Note: Proposed (MaskTRN + DaU + PaR)

Calibration curves of different models.

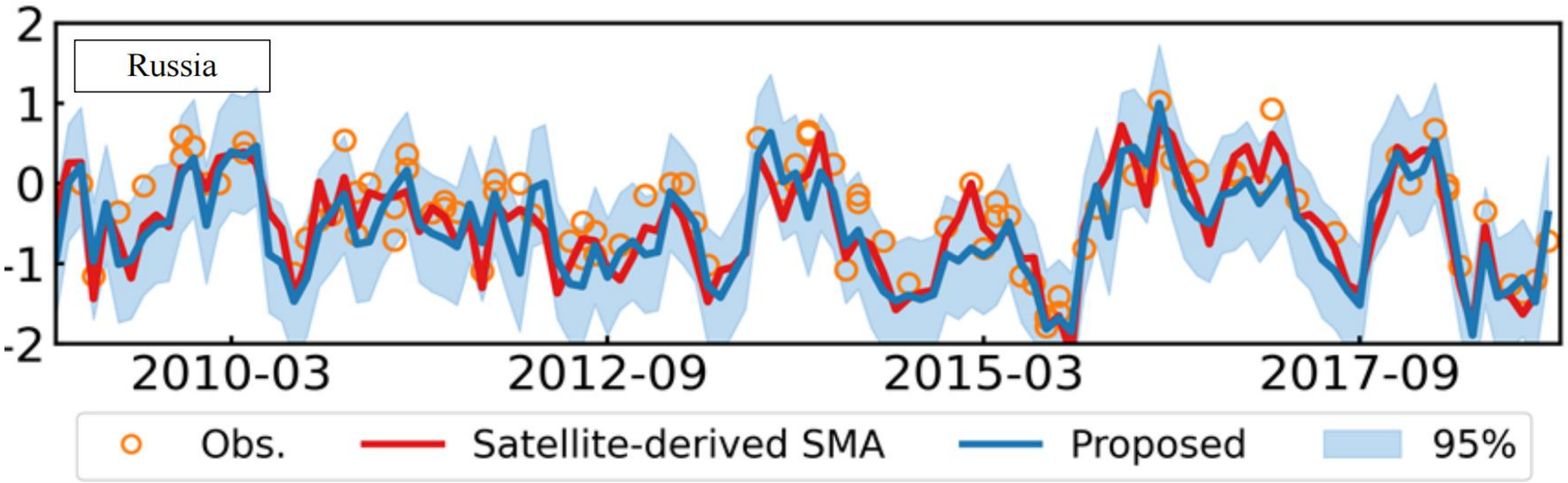
3.7 Qualitative Results



Note: Proposed (MaskTRN + DaU + PaR)

LSM-based: spatially aggregated SMA and uncertainty signals

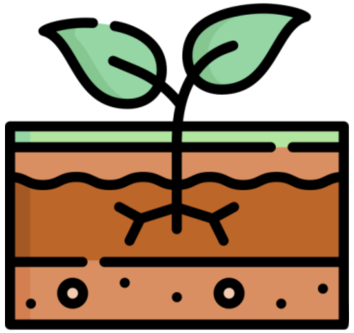
3.7 Qualitative Results



Note: Proposed (MaskTRN + DaU + PaR)

Satellite-derived: spatially aggregated SMA and uncertainty signals

3.8 Summary



- DaU improves model generalization to future climate
- PaR enforces consistency with local processes
- Reliable predictions with accurate uncertainty and consistency

Mengxue Zhang, Miguel-Ángel Fernández-Torres, and Gustau Camps-Valls, Long-Term Soil Moisture Drought Forecasts Using a Robust Deep Learning Framework, *Under Review, 2025*

Conclusion

Contribution of this Thesis

- Development of a hybrid probabilistic deep learning model for drought detection.

- Calibration and uncertainty framework for probabilistic drought prediction.

- Robust DL framework for long-term soil moisture drought forecasting.

Significance:
This thesis demonstrates that integrating domain knowledge, probabilistic inference, and deep learning can substantially improve drought monitoring and forecasting

Related Outcomes and Projects

1. **Mengxue Zhang**, Miguel-Ángel Fernández-Torres, and Gustau Camps-Valls. "Hybrid recurrent neural network for drought monitoring." In Proc. NeurIPS, pp. 1-9. 2022.

2. **Mengxue Zhang**, Kai-Hendrik Cohrs, Miguel-Ángel Fernández-Torres, Esther Rodrigo-Bonet, and Gustau Camps-Valls. " The Risks of Imperfect Knowledge: Reliability Trade-Offs in Physics-Aware Neural Networks for Climate Projections. " International Conference on Artificial Neural Networks, 2026, accepted.

3. Gonzalez-Calabuig, Maria, Jordi Cortés-Andrés, Tristan Keith Ellis Williams, **Mengxue Zhang**, Oscar Jose Pellicer-Valero, Miguel-Angel Fernandez-Torres, and Gustau Camps-Valls. "The AIDE toolbox: Artificial intelligence for disentangling extreme events [software and data sets]." IEEE Geoscience and Remote Sensing Magazine 12, no. 2 (2024): 113-118.

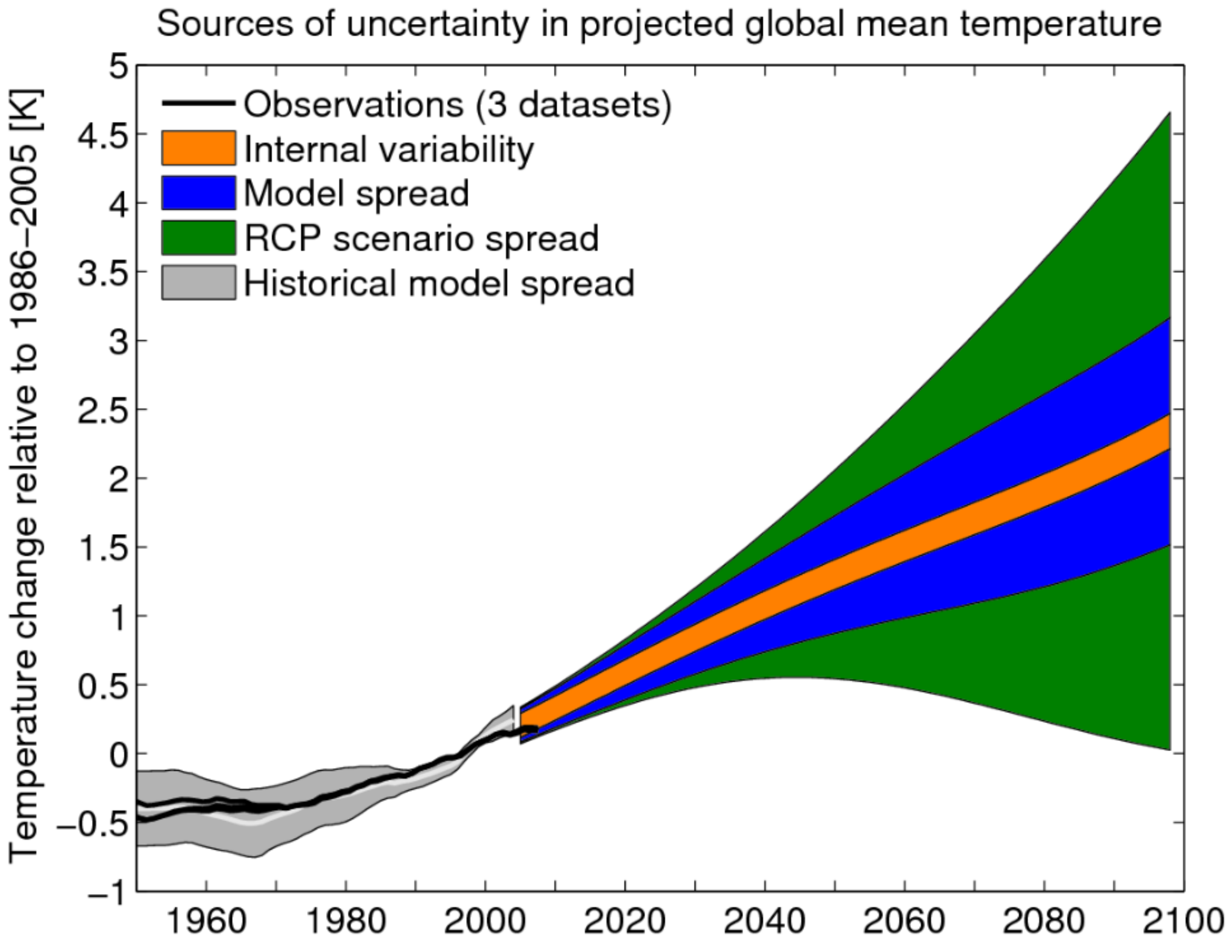
Limitations of this Thesis

- Domain knowledge priors may introduce unrealistic biase

- Estimated uncertainty may be physically implausible

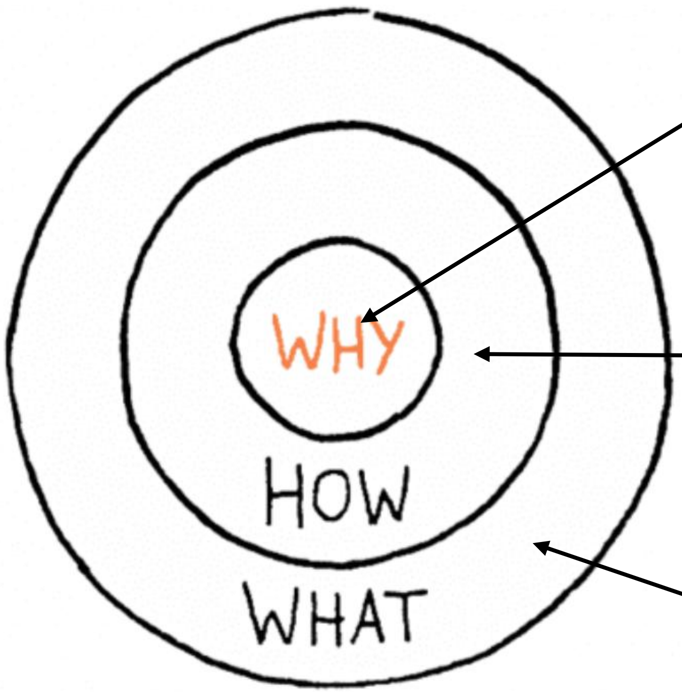
- Calibration relies on assumptions of shifts

Future Work



Uncertainty-aware domain knowledge prior

Future Work



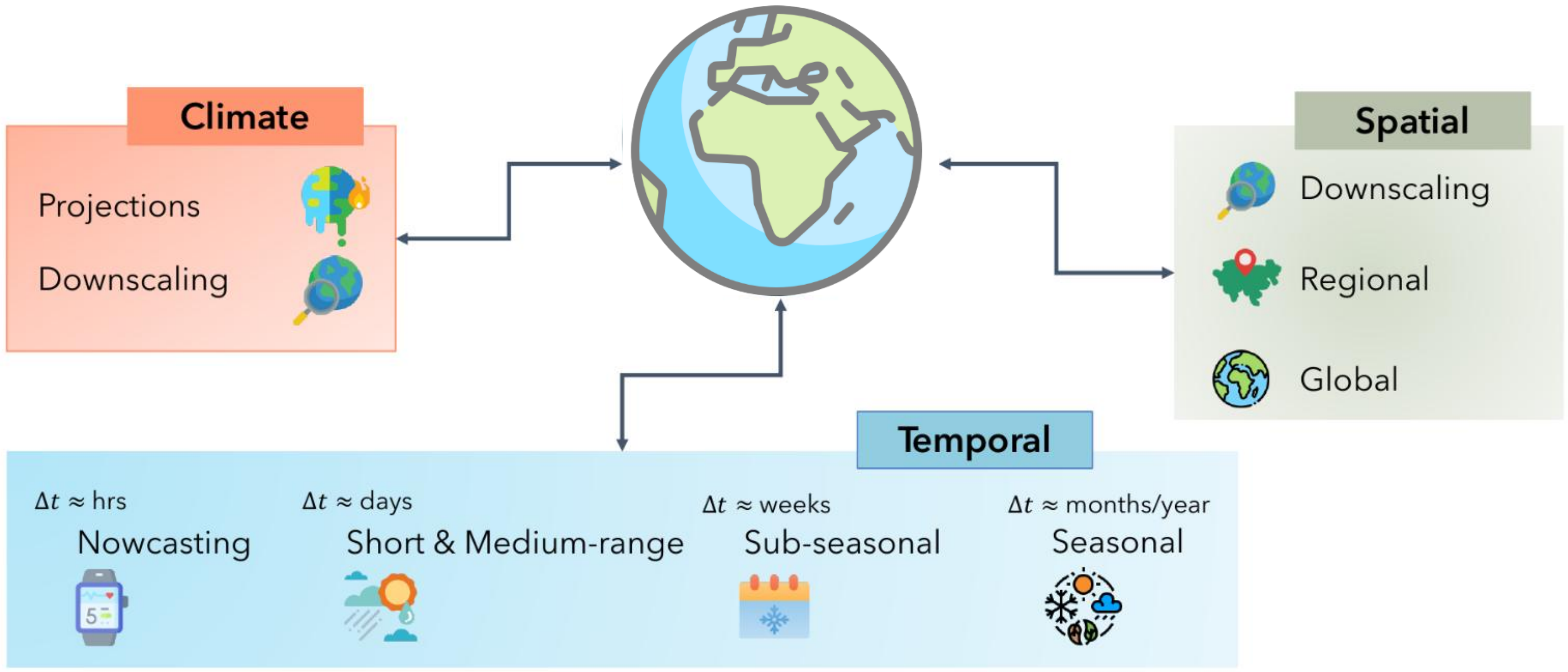
parametric and structural misspecification, sensitivity, chaos.

spatial, temporal, component, propagation and interaction.

epistemic, aleatoric uncertainty, distribution shifts.

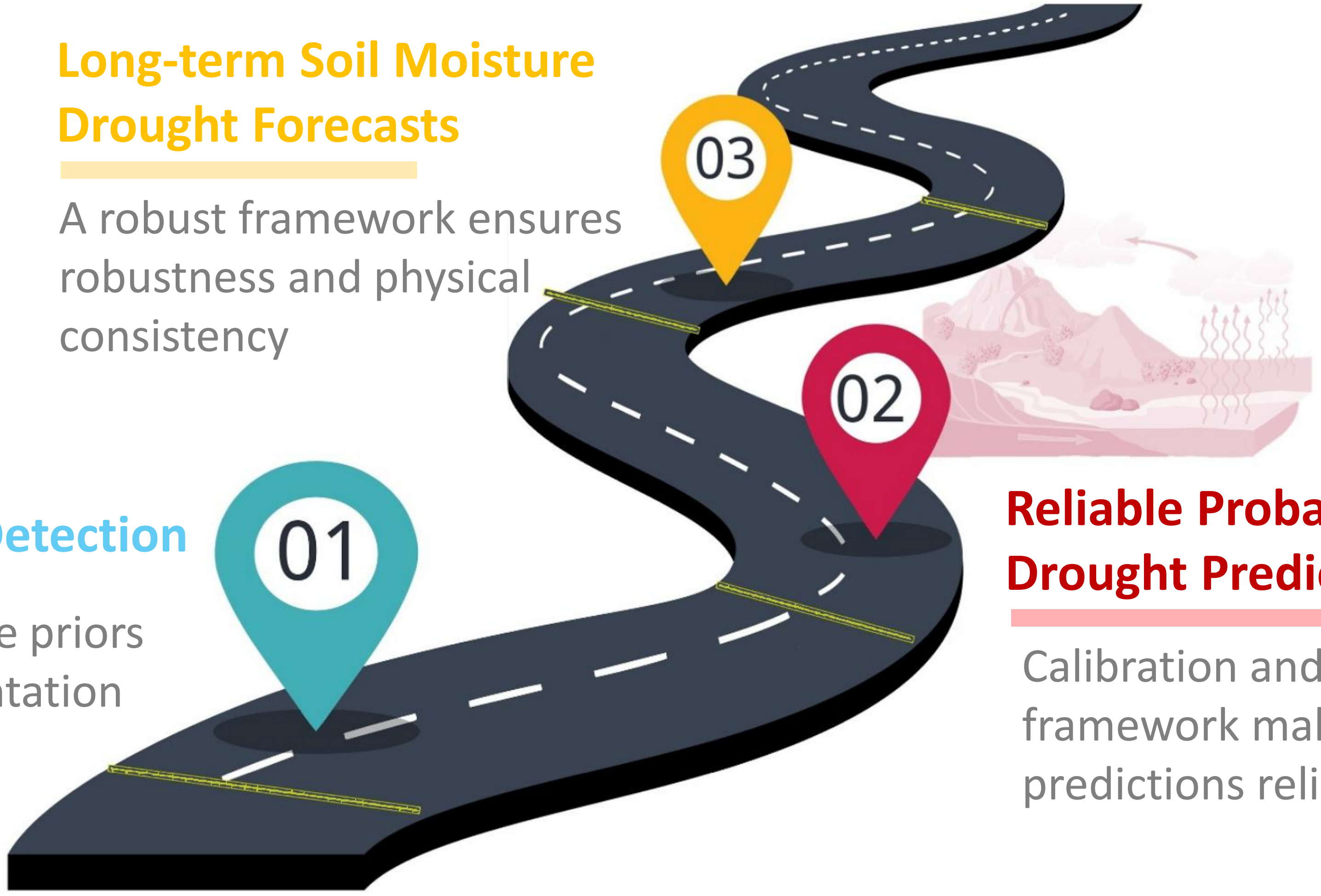
Physical consistency and causality of uncertainty

Future Work



Knowledge-informed calibration for GeoFM

Final Remarks



Drought Event Detection

Domain knowledge priors improves representation and generalization

Reliable Probabilistic Drought Prediction

Calibration and uncertainty framework makes predictions reliable

Long-term Soil Moisture Drought Forecasts

A robust framework ensures robustness and physical consistency

Acknowledge

The author acknowledges funding from the China Scholarship Council (CSC, No. 202106710031). This thesis is further supported by European Research Council (ERC) Synergy Grant USMILE (855187), Horizon 2020 projects XAIDA (101003469) and AI4PEX (101137682). Computational resources were provided by the Jülich Supercomputing Centre (PRACE-DEV-2022D01-048) and Artemisa, with additional technical support from the IFIC, CSIC-UV.