

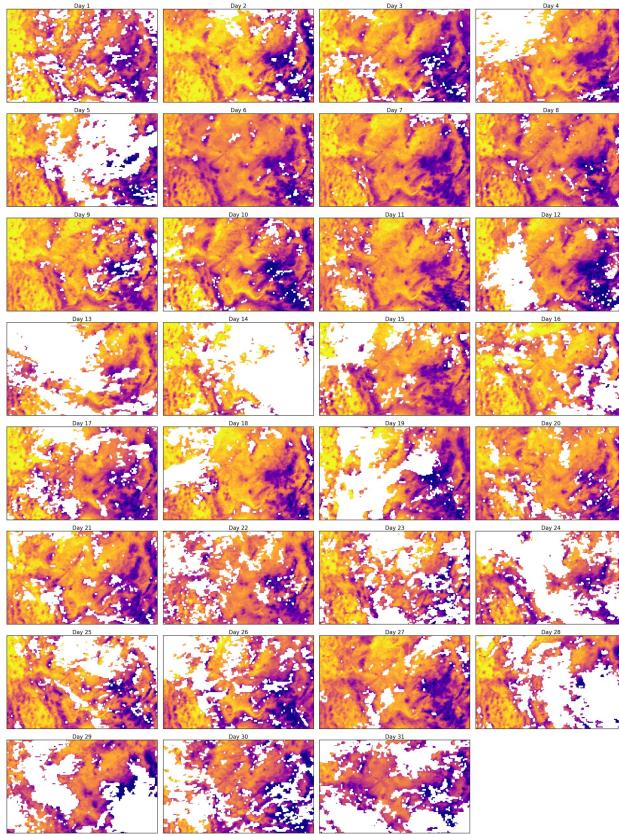
Spatio-Temporal Modeling of Land Surface Temperature: A Comparison of IDW, Random Forest, and Sparse Gaussian Processes

Course project Presentation

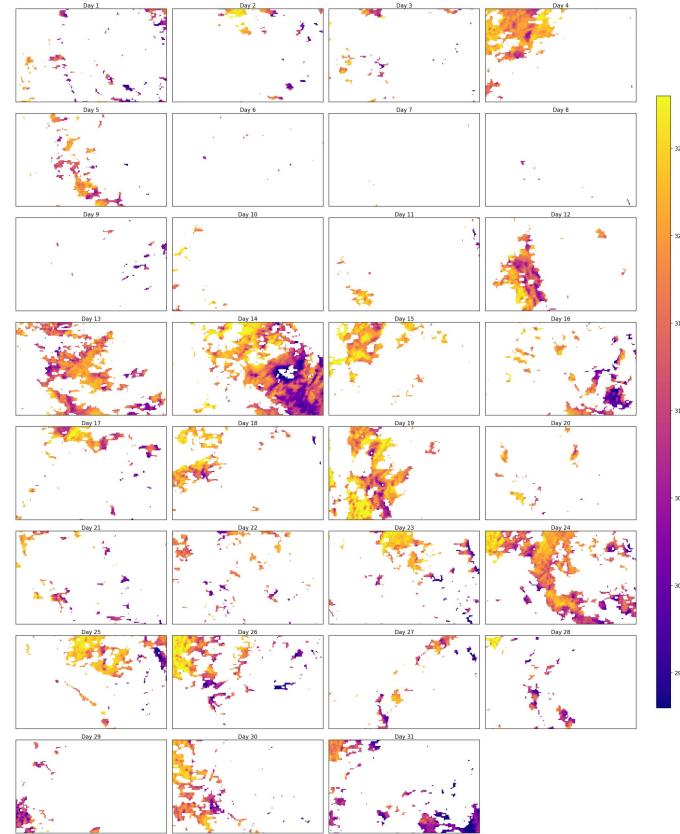
Presenter: Xiaoning Sun

Student number: 261162300

1. Background and Introduction



Training set



Test set

Objective of the project:

Develop probabilistic interpolation models for land surface temperature (LST) data collected from the MODIS satellite platform.

Dataset: $100 \times 200 \times 31$ (latitude \times longitude \times days)

- Latitude: 35° to 40° N (100 grid points)
- Longitude: -115° to 105° W (200 grid points)
- Time: 31 consecutive days
- 494,762 training points
- 85,942 test points
- No overlap between training and test sets.

2. Project Questions & Objectives

Q1. Model Performance

How well do models reconstruct MODIS LST?

IDW, Random Forest (RF), Sparse Variational Gaussian Process (SVGP)



Q2. Kernel Design

Which covariance kernels best capture LST structure in (lon, lat, time)?

Linear, RBF, Matern 3/2, Matern 5/2, RQ-Like, RBF+Linear, Periodic \times Time, Space \times Time



Q3. Role of Additional information (Elevation)

How does elevation improve LST interpolation? What is the best integration strategy?

Strategy 1: SVGP+Elevation as input feature

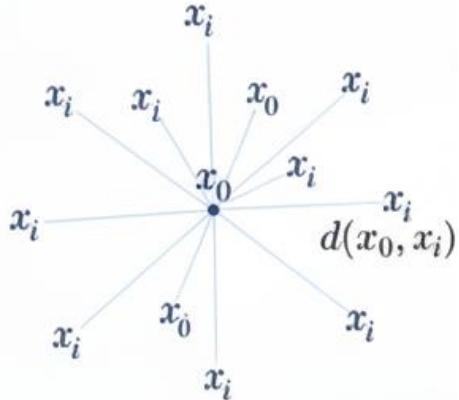
Strategy 2: Multi-output SVGP (Elevation + Temperature)

ICM (Intrinsic Coregionalization Model)

3. Model choice

3.1 Baseline Models: IDW & Random Forest

Model 1: Inverse Distance Weighting (IDW)



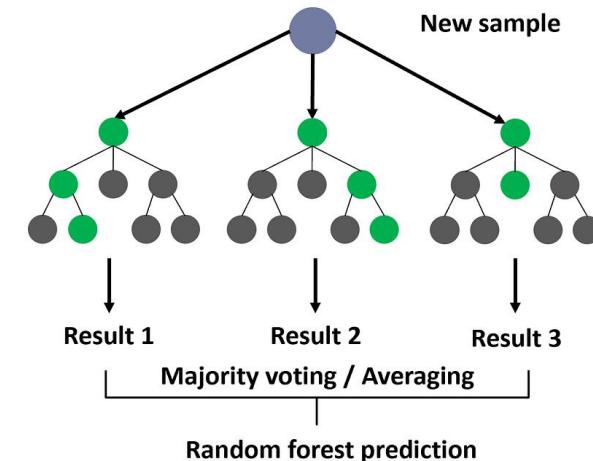
$$\hat{y}(x_0) = \frac{\sum_i w_i y_i}{\sum_i w_i}$$

$$\text{where } w_i = \frac{1}{d(x_0, x_i)^p}.$$

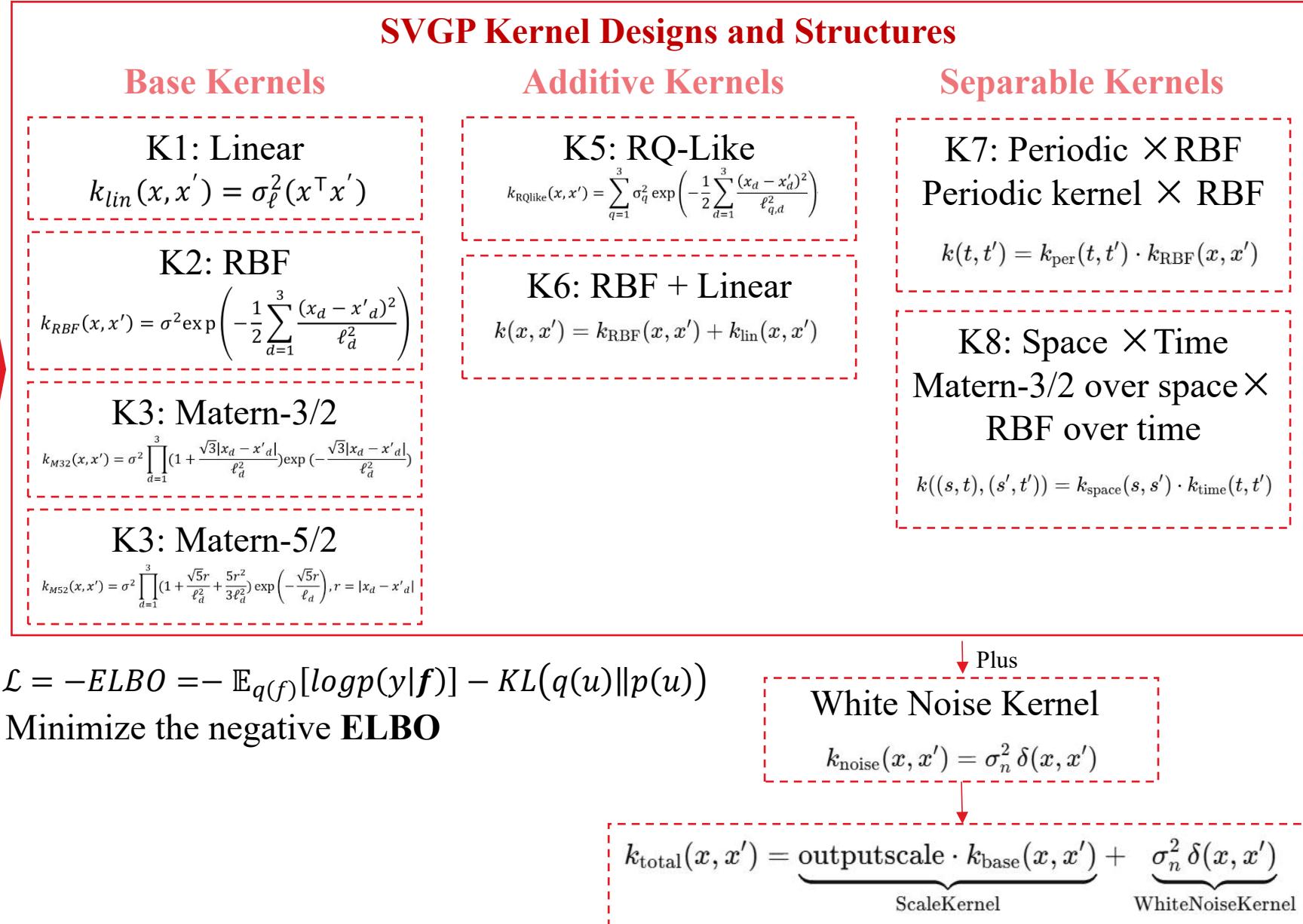
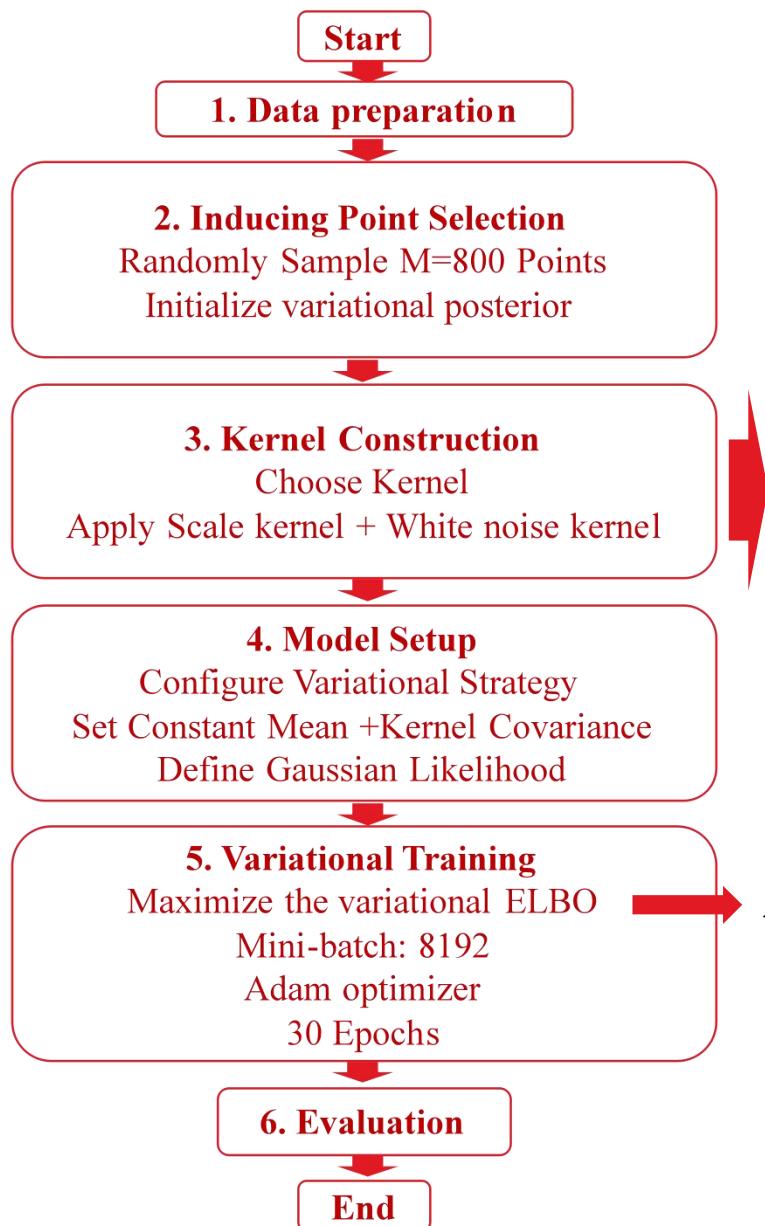
- Deterministic spatial interpolation
- Local weighted averaging of nearby observations
- Weights decay with distance
- Applied independently for each day → no temporal modeling
- No uncertainty quantification
- Simple, fast, and widely used as a baseline method

Model 2: Random forest model

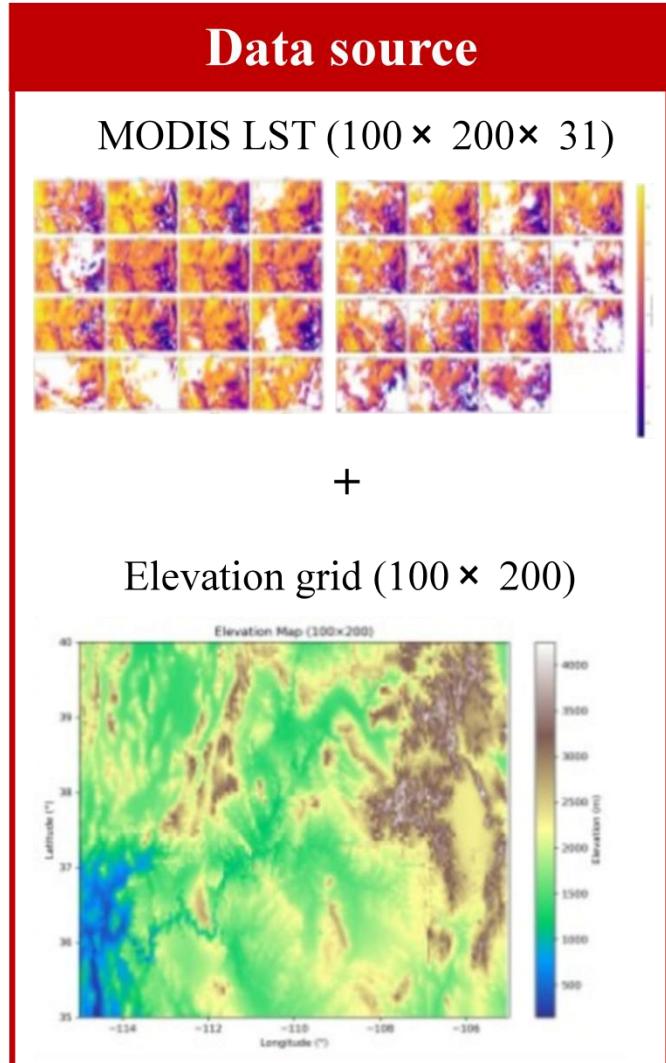
- Ensemble of many decision trees (bagging)
- Captures nonlinear relationships effectively
- Treats (row,col,time) as regular features
- Robust to noise and requires little tuning
- No uncertainty quantification
- Not a spatiotemporal model → may produce non-smooth spatial patterns



3.2 Sparse Variational Gaussian Process (SVGP)



3.3 SVGP with elevation



Processing

- Reshape LST tensor to training samples
- Split training vs test data
- Normalize inputs (row, col, day)
- Align elevation grid with spatial coordinates
- Produce training X, y

Strategy 1 : SVGP + Elevation as Input Feature

- Input: $X = [\text{row}, \text{col}, \text{day}, \text{elevation}]$
- Kernel design:
- $k_{\text{total}} = k_{\text{spacetime}} + k_{\text{elev}}$
- (RBF / Matérn5/2 / Space \times Time)
- Sparse Variational GP (SVGP)
- Complexity $O(NM^2)$

Strategy 2 — Multi-output SVGP

- Input: $X = [\text{row}, \text{col}, \text{day}]$
- Outputs:
 $y_1 = \text{LST} (\text{temperature}), y_2 = \text{elevation}$
- Shared inducing points Z
- Shared kernel hyperparameters
- (lengthscale & outputscale for RBF/Matérn)
- Joint ELBO = ELBO_temp + ELBO_elev

Data source: OpenTopography- NASADEM Global Digital Elevation Model

4. Metrics Design and computation cost

Error-based Metrics

[1] Root Mean Squared Error (RMSE)

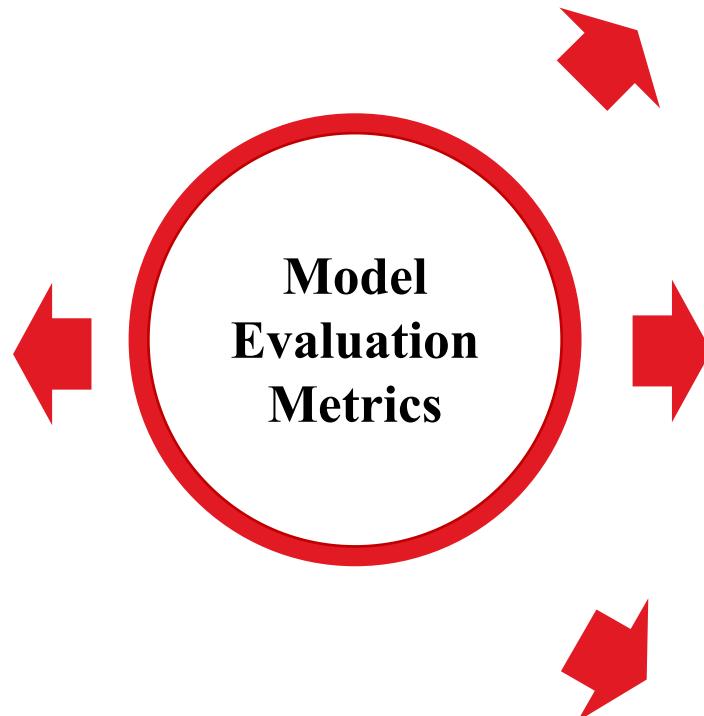
$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

[2] Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

[3] Coefficient of Determination

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$



Probabilistic Metrics

[4] Continuous Ranked Probability Score (CRPS)

$$\text{CRPS}(y, \mu, \sigma) = \sigma [z(2\Phi(z) - 1) + 2\phi(z)] - \frac{\sigma}{\sqrt{\pi}}$$

Interval Metrics

[5] Prediction Interval Coverage Probability (PICP)

$$\text{PICP} = \frac{1}{N} \sum_{i=1}^N I(y_i \in [L_i, U_i])$$

[6] Mean Prediction Interval Width (MPIW)

$$\text{MPIW} = \frac{1}{n} \sum_{i=1}^n (U_i - L_i)$$

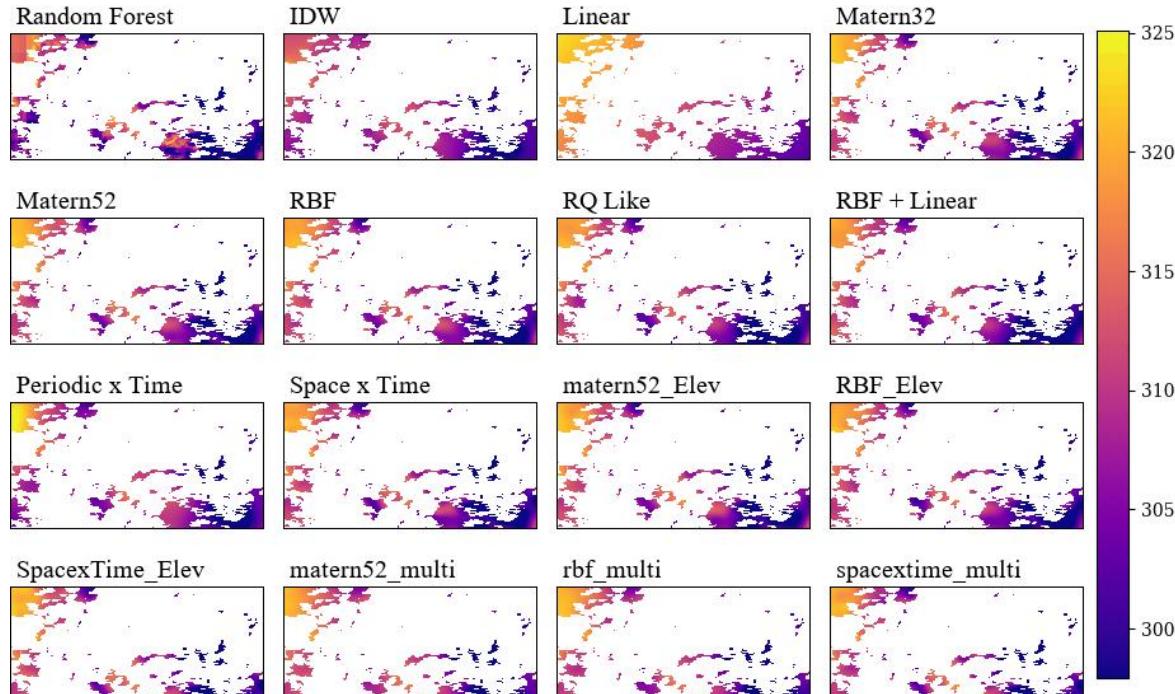
Efficiency

[7] Training Time

5. Results

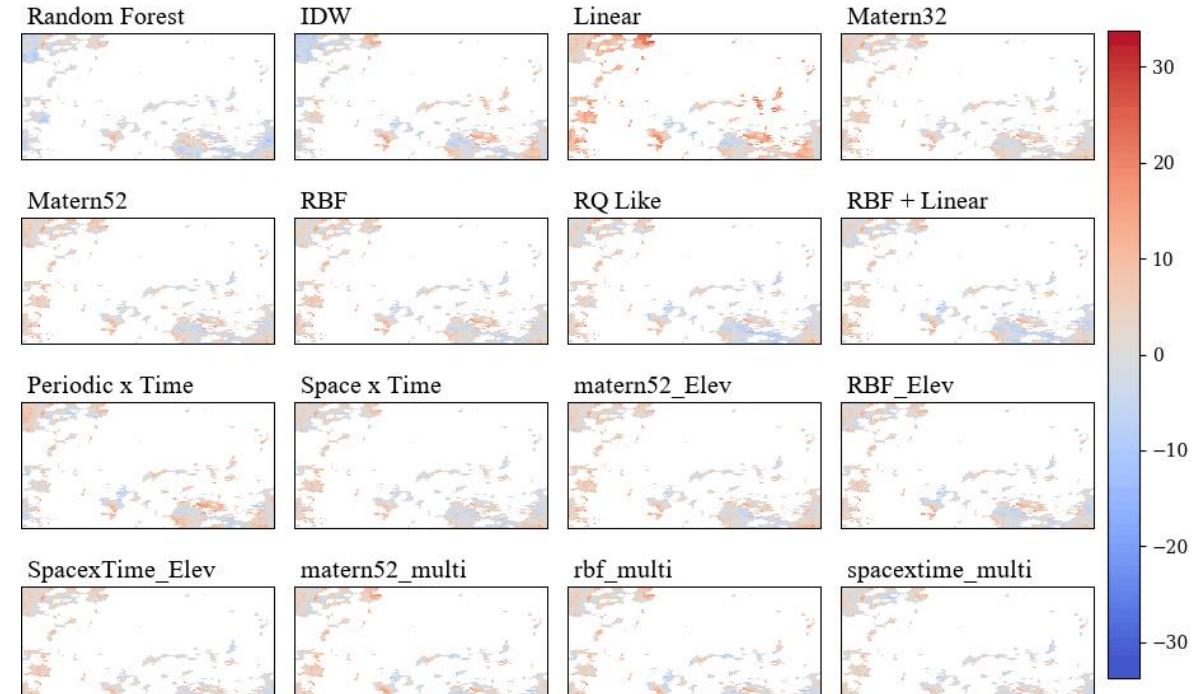
5.1 Results Overview

Prediction maps



Day 31

Error maps



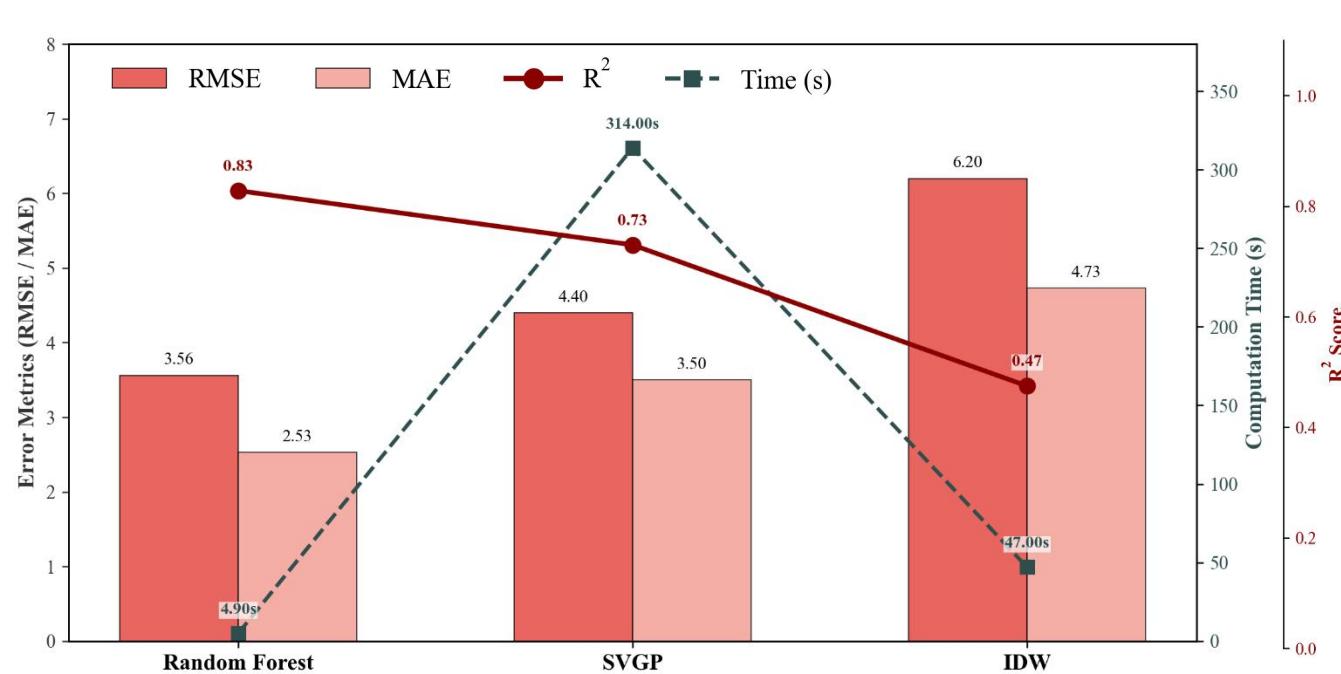
Day 31

5.2 Metrics Results

Model	Kernel	RMSE (K)	MAE	R2	CRPS	PICP	MPIW	Time(s)
IDW	/	6.20	4.73	0.47	/	/	/	47.09
Random Forest	/	3.56	2.53	0.83	/	/	/	4.90
SVGP	Linear	7.39	5.96	0.25	4.17	0.95	28.02	263.32
	RBF	4.73	3.72	0.69	2.65	0.96	19.23	282.76
	Matern3/2	4.60	3.59	0.71	2.58	0.96	19.41	311.14
	Matern5/2	4.59	3.58	0.71	2.56	0.95	18.79	303.28
	rq_like(Mixture of RBFs)	4.99	3.91	0.66	2.79	0.95	20.13	329.13
	RBF + Linear	5.03	3.91	0.65	2.80	0.95	20.26	300.58
	Periodic \times time	5.47	4.24	0.59	3.04	0.94	20.39	324.39
	Space \times time Matern3/2(x, y) \times RBF(t)	4.40	3.43	0.73	2.46	0.96	18.27	314.06
SVGP+ Elevation (input feature)	RBF_Elev	4.76	3.73	0.69	2.67	0.96	19.52	398.81
	Matern5/2_Elev	4.62	3.62	0.71	2.59	0.96	19.13	381.83
	Space \times time_Elev	4.44	3.46	0.73	2.48	0.96	18.45	347.45
Multi-output SVGP	RBF_multi	5.34	4.18	0.61	2.98	0.94	20.64	690.73
	Matern5/2_multi	5.26	4.12	0.62	2.94	0.94	20.36	592.98
	Space \times time_multi	4.44	3.47	0.73	2.48	0.95	18.23	630.96

5.3 Comparaison of different models

Model	Kernel	RMSE	MAE	R2	CRPS	PICP	MPIW	Time(s)
IDW	/	6.20	4.73	0.47	/	/	/	47.09
Random Forest	/	3.56	2.53	0.83	/	/	/	4.90
SVGP	Space×time	4.40	3.43	0.73	2.46	0.96	18.27	314.06



1. Random Forest (Best Accuracy)

Lowest RMSE, Highest R², Fastest model

Limitation: No uncertainty estimates

2. SVGP (Best Probabilistic Performance)

Good accuracy

Provides **well-calibrated uncertainty**

Trade-off: Highest computational cost (314 s)

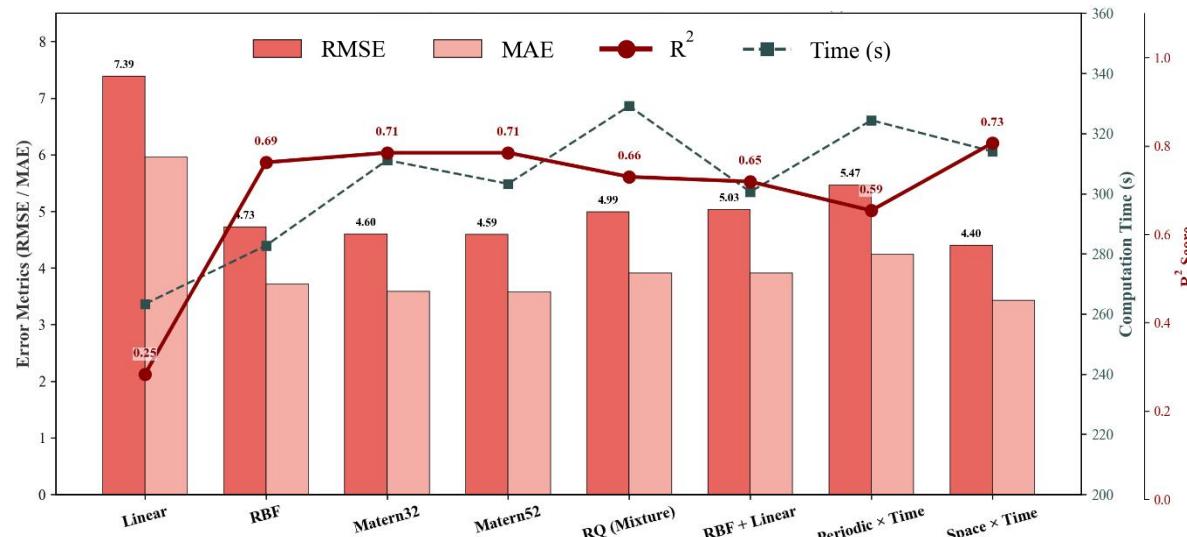
3. IDW (Weakest Performance)

Highest error, Lowest R², No uncertainty estimation

Moderate runtime (47 s)

5.4.1 Comparaison of different kernels

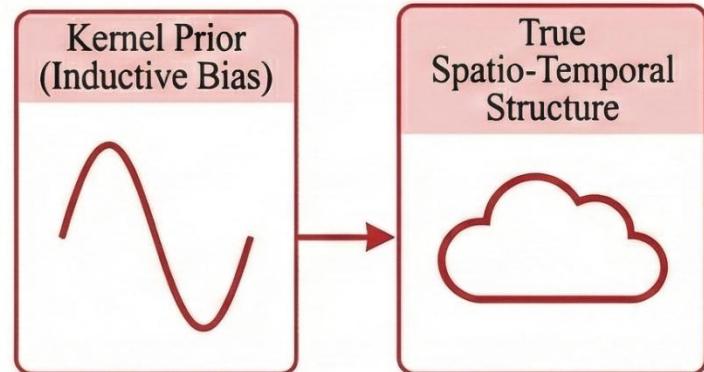
Model	Kernel	RMSE	MAE	R2	CRPS	PICP	MPIW	Time(s)
SVGP	Linear	7.39	5.96	0.25	4.17	0.95	28.02	263.32
	RBF	4.73	3.72	0.69	2.65	0.96	19.23	282.76
	Matern3/2	4.60	3.59	0.71	2.58	0.96	19.41	311.14
	Matern5/2	4.59	3.58	0.71	2.56	0.95	18.79	303.28
	rq_like(Mixture of RBFs)	4.99	3.91	0.66	2.79	0.95	20.13	329.13
	RBF + Linear	5.03	3.91	0.65	2.80	0.95	20.26	300.58
	Periodic \times time	5.47	4.24	0.59	3.04	0.94	20.39	324.39
	Space \times time	4.40	3.43	0.73	2.46	0.96	18.27	314.06



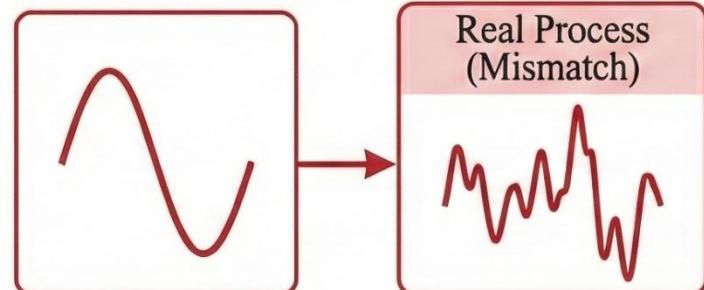
- 1. Space–Time kernel**
→ most effective and best calibrated
- 2. Matern kernels**
→ reliable, strong alternatives
- 3. RBF**
→ moderate models
- 4. Composite and Linear kernels**
→ significantly less effective

5.4.2 Why Different Kernels Perform Differently

1. Kernel Performance Depends on Prior-Data Match

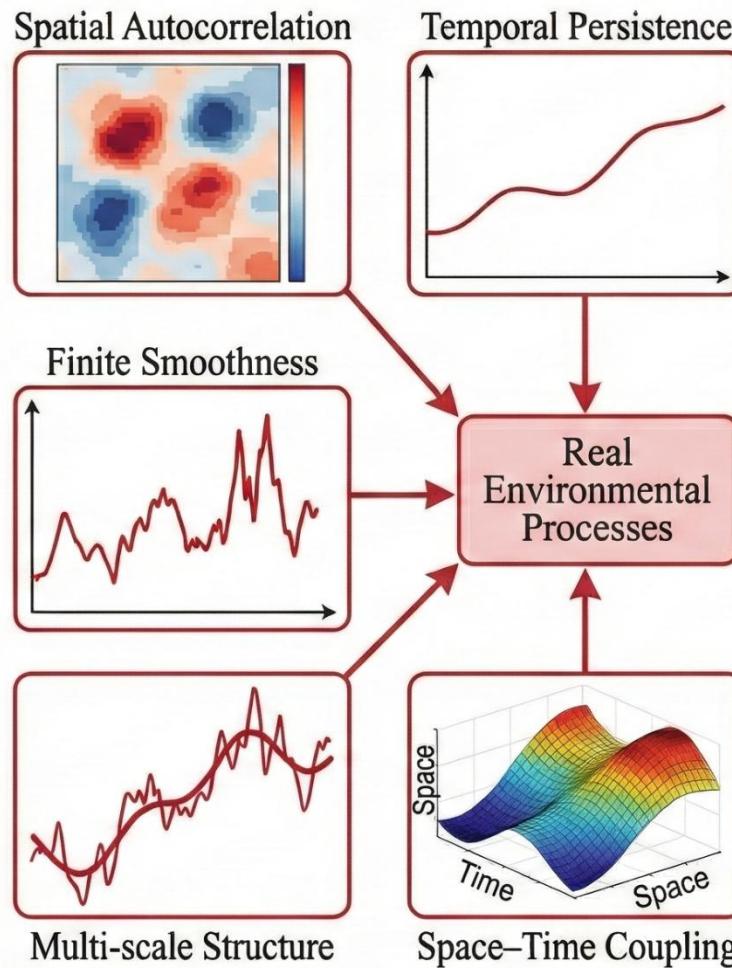


Alignment → Improved Performance

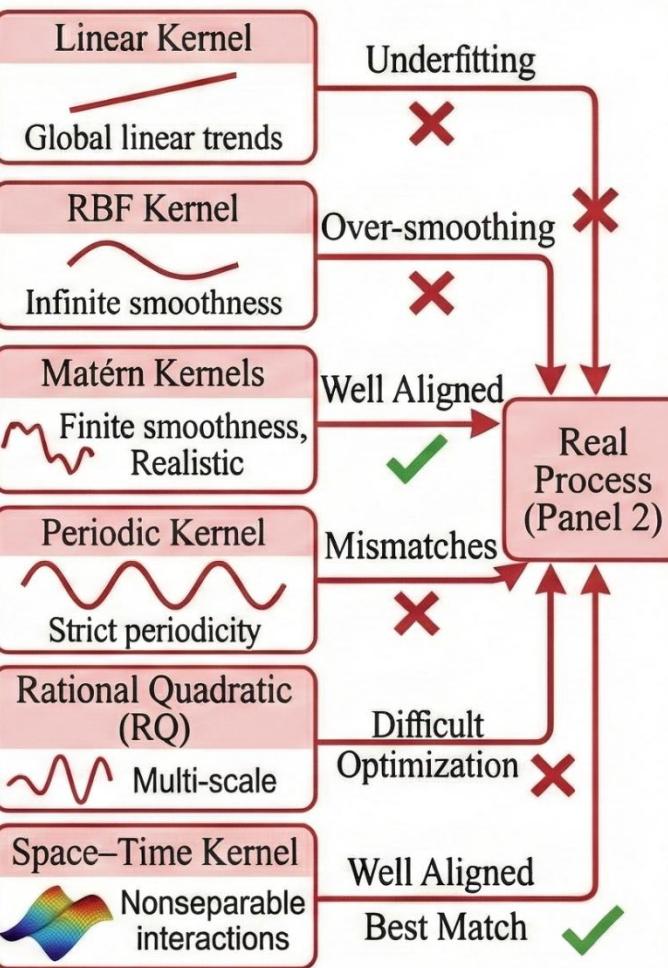


Mismatch → Systematic Bias,
Model Misspecification

2. Real Environmental Process Properties



3. Different Kernels Correspond to Different RKHS Function Spaces



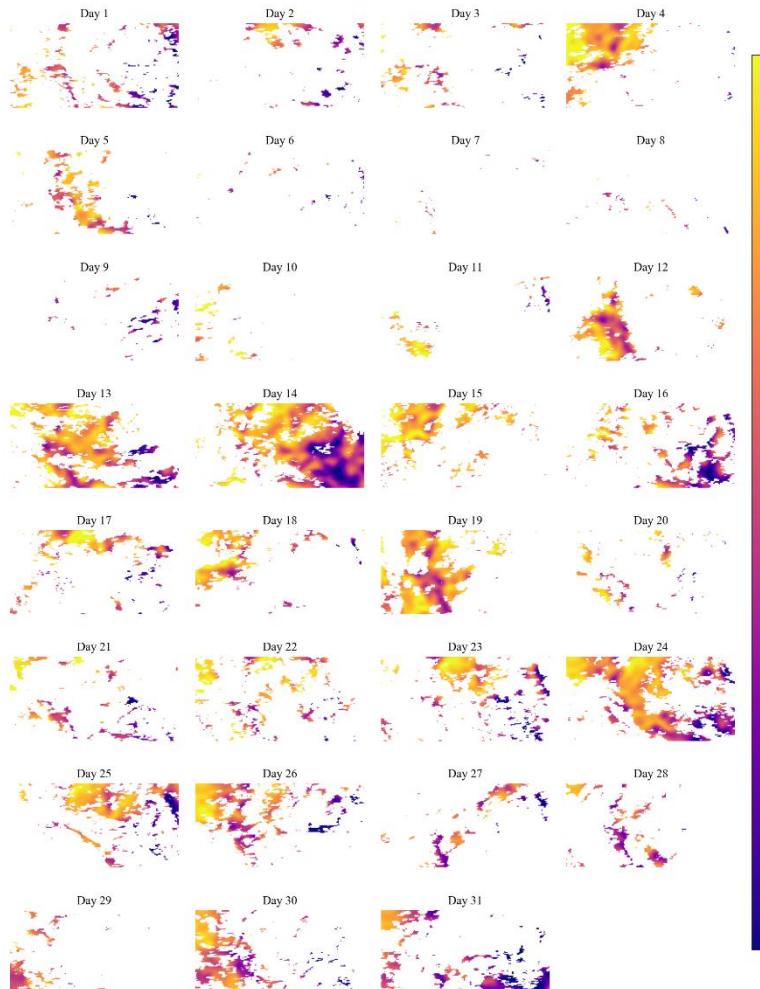
1. Kernel Prior–Data Match

2. Real Environmental Process Properties

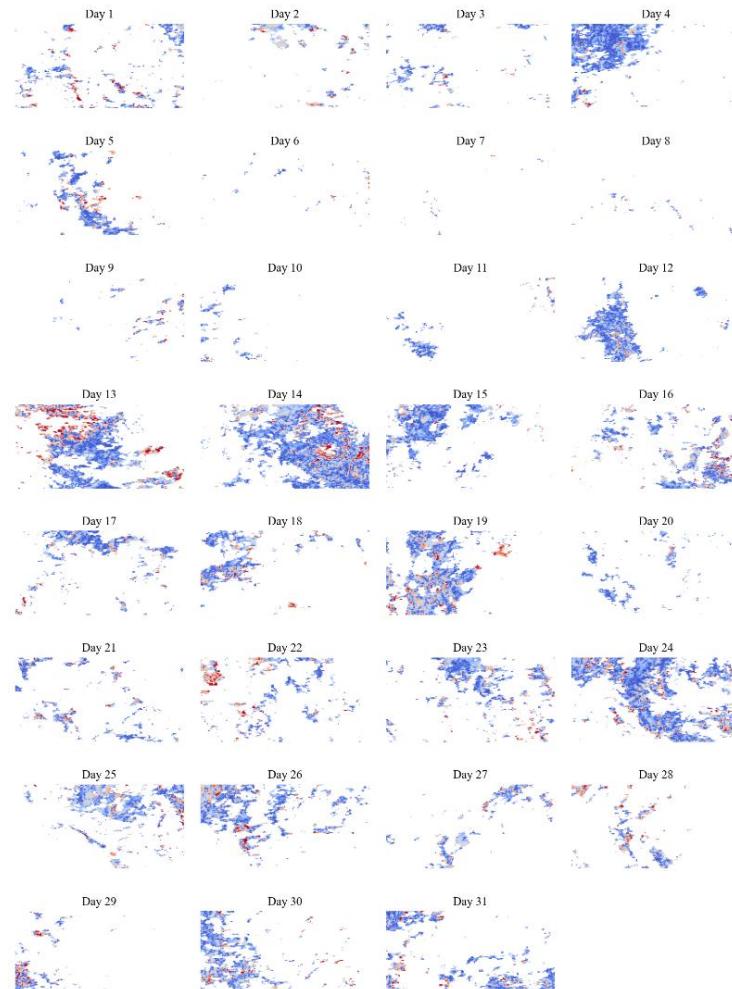
3. Kernel RKHS Function Spaces

5.4.3 Spatio-Temporal Results of the Space × Time Kernel

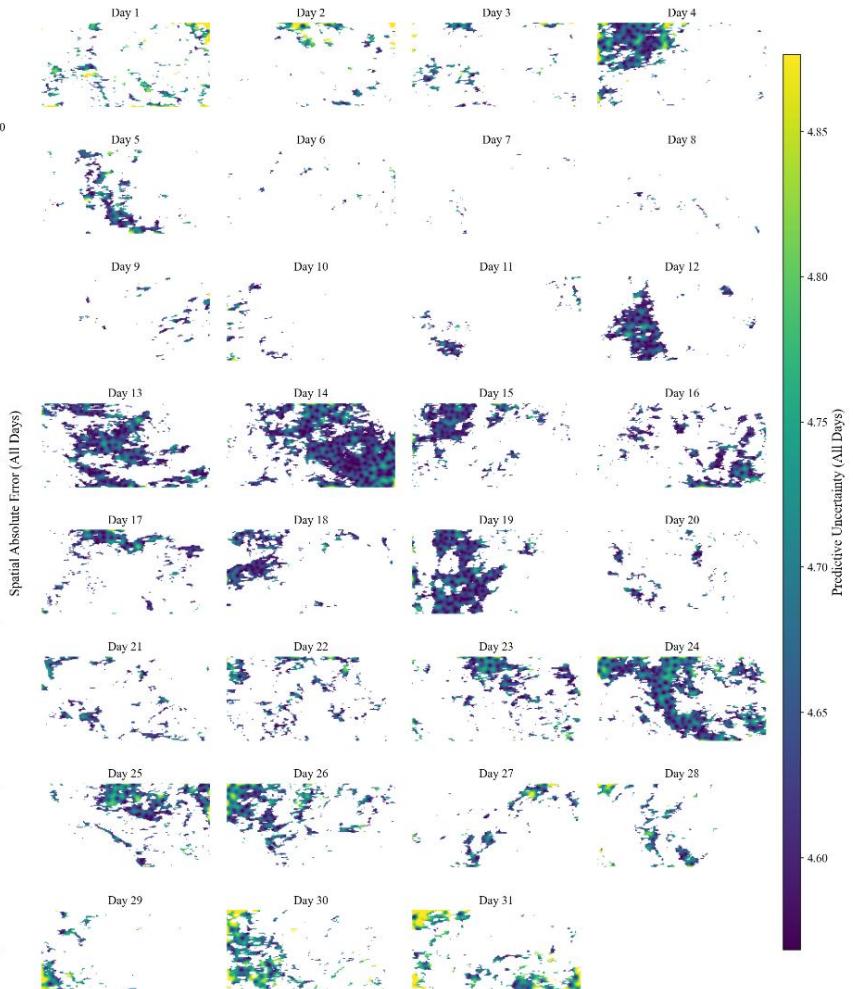
Space × Time Kernel: Prediction



Space × Time Kernel: Spatial Absolute Error



Space × Time Kernel: Predictive Uncertainty



5.4.4 Interpreting the Space × Time Kernel Performance

1. Prediction Maps: Strong Spatio-Temporal Coherence

- Smooth spatial gradients and temporally consistent evolution patterns across all 31 days.
- Capture nonseparable spatio-temporal dependencies (broad patterns & localized transitions)

2. Spatial Error: Concentrates in High-Variability Regions

- Errors remain mostly low, consistent with strong RMSE/R² performance.
- Discrepancies localize in areas with:
 - Rapid temporal changes
 - Strong spatial gradients
 - Dynamic boundaries

3. Predictive Uncertainty: Well-Calibrated & Interpretable Patterns

Spatially structured (Not noise)

- Forms coherent patterns representing systematic epistemic uncertainty

Physically Meaningful:

- Uncertainty increases in:
 - Data-sparse regions
 - Complex dynamics or rapid changes

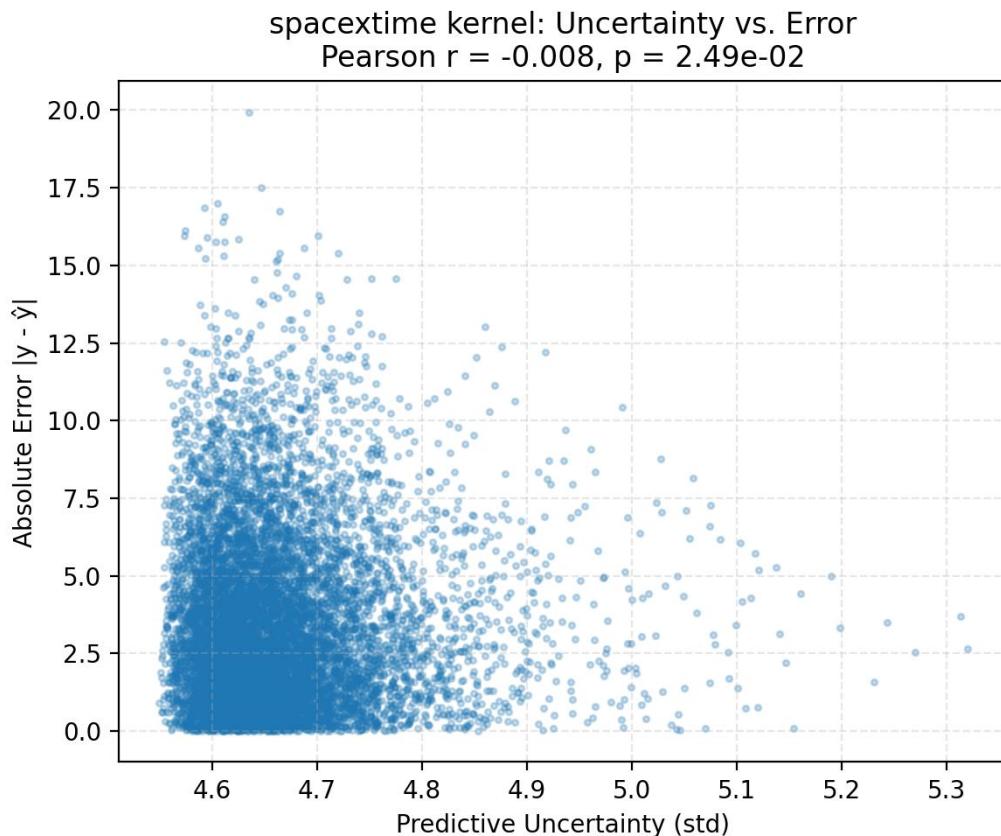
Excellent Calibration

High sharpness (narrow intervals) + correct coverage (PICP=0.96) indicates an exceptionally well-calibrated GP.

4. Integrated Interpretation: Model Validation

Accurate mean predictions, spatially coherent and interpretable residual patterns, well-structured, meaningful uncertainty estimates aligned with both data density and physical complexity.

5.4.5 Relationship Between Predictive Uncertainty and Error



Key Finding: Predictive uncertainty (std) from the Space–Time SVGP is almost uncorrelated with actual prediction error.

Why This Happens:

- Variational Posterior Variance Collapse
 - predictive std falls within a very narrow range
- Homoscedastic Gaussian Likelihood
 - predictive uncertainty is largely dominated by data geometry (distance to training samples) rather than the intrinsic difficulty of the prediction task.
- Strong Smoothness Imposed by the Space–Time Kernel
- Environmental Data Are Highly Smooth

Implications:

- Uncertainty Does Not Track Error
- Globally Calibrated but Not Locally Calibrated
- Uncertainty Reflects Data Density, Not Epistemic Risk

5.5.1 Effect of Incorporating Elevation into the SVGP Model

Model	Kernel	RMSE	MAE	R2	CRPS	PICP	MPIW	Time(s)
SVGP	RBF	4.73	3.72	0.69	2.65	0.96	19.23	282.76
	Matern5/2	4.59	3.58	0.71	2.56	0.95	18.79	303.28
	Space×time	4.40	3.43	0.73	2.46	0.96	18.27	314.06
SVGP+ Elevation (input feature)	RBF_Elev	4.76	3.73	0.69	2.67	0.96	19.52	398.81
	Matern5/2_Elev	4.62	3.62	0.71	2.59	0.96	19.13	381.83
	Space×time_Elev	4.44	3.46	0.73	2.48	0.96	18.45	347.45
Multi-output SVGP	RBF_multi	5.34	4.18	0.61	2.98	0.94	20.64	690.73
	Matern5/2_multi	5.26	4.12	0.62	2.94	0.94	20.36	592.98
	Space×time_multi	4.44	3.47	0.73	2.48	0.95	18.23	630.96

- Adding Elevation as an input yields only marginal gains**

Minimal improvement (e.g., $4.40 \rightarrow 4.44$)

20–30% higher computation

Indicates elevation effect is mostly absorbed by spatial coordinates

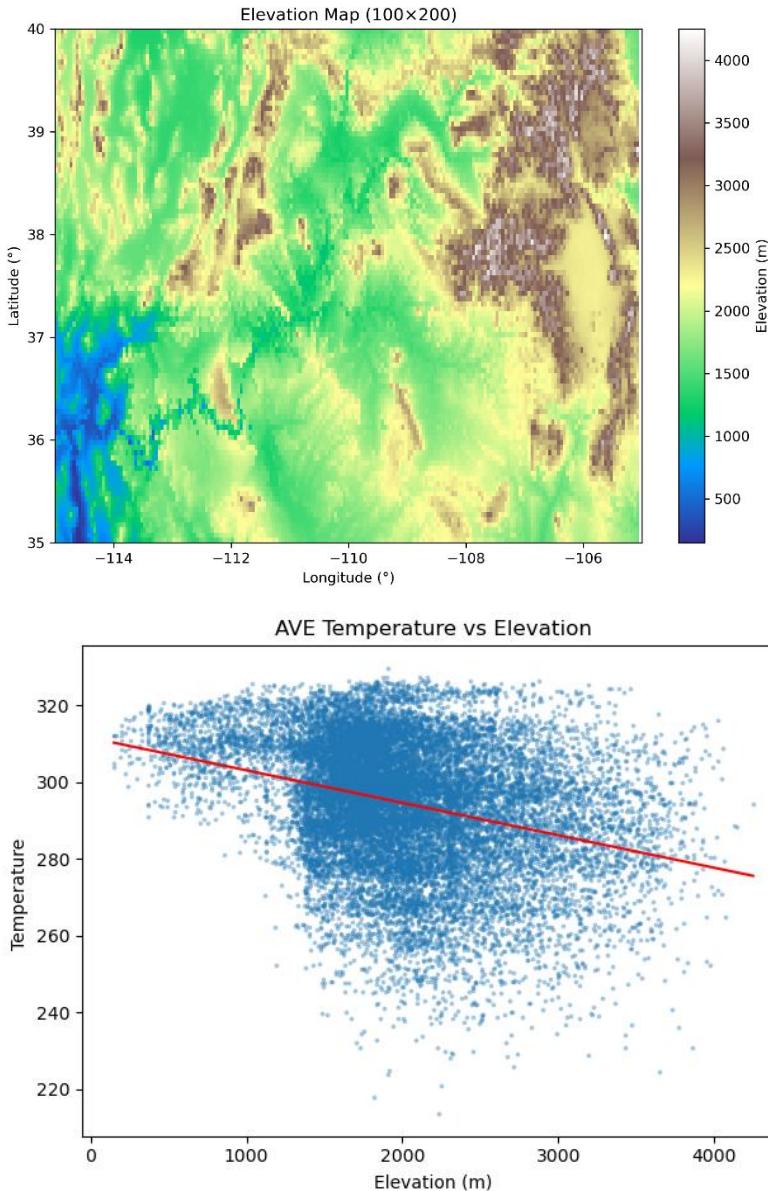
- Multi-output SVGP increases cost but not accuracy**

Training time nearly doubles

Very similar RMSE and R^2

Cross-output correlation too weak to offer benefit

5.5.2 Why Elevation Fails to Improve Prediction Accuracy



Insights from Temperature–Elevation Scatter Plot

- Clear global negative trend → elevation matters
- But large vertical spread → strongly nonstationary & heteroscedastic effect
- Elevation strongly collinear with spatial coordinates
- Sparse high-elevation samples

Why Elevation Adds Little Benefit?

- Elevation–temperature relationship is nonstationary
- Spatial coordinates already encode major elevation trends
- Uneven elevation distribution impairs uncertainty learning
- Model ends up capturing only a weak global trend

Why Multi-output SVGP Fails to Improve Performance?

- Cross-output correlation is weak → limited information sharing
- Shared inducing points lead to underfitting each output
- Increased complexity does not increase representation power
- High computational cost with minimal benefit

6. Conclusion

1. SVGP provides a principled probabilistic framework

- Balances predictive accuracy and uncertainty quantification.
- More suitable for environmental modeling than deterministic methods (RF/IDW).

2. Space × Time kernel achieves the best performance

- Captures spatial autocorrelation and temporal continuity.
- Best results in RMSE, spatial coherence, and probabilistic metrics (CRPS/PICP).

3. Elevation contributes limited improvement

- Temperature–elevation relationship is highly non-stationary.
- Latitude/longitude already encode most terrain structure → marginal gain.
- Multi-output GP increases computational cost with little benefit.

4. Key limitation: local uncertainty does not track error

- Variational inference leads to variance collapse and narrow predictive std.
- Global calibration is good, but uncertainty \neq local error.

7. Future work

1. Improved uncertainty modeling

- Heteroscedastic or non-Gaussian likelihoods for adaptive local uncertainty.

2. More expressive kernels

- Deep Kernel Learning (DKL), non-stationary kernels, spatially varying length-scales.

3. Enhanced calibration

- Conformal Prediction for locally reliable prediction intervals.

CIVE 650 – Spatiotemporal Data Mining

Course project presentation

Thanks for your listening!

Presenter: Xiaoning Sun
Student number: 261162300

[Github link: xiaoning-droid/data_mining_project](#)