SimFair: Physics-Guided Fairness-Aware Learning with Simulation Model

Zhihao Wang[1], Yiqun Xie[1*], Zhili Li[1], Xiaowei Jia[2], Zhe Jiang[3], Aolin Jia[1], Shuo Xu[1]

[1]University of Maryland, [2]University of Pittsburgh, [3]University of Florida

{zhwang1, xie, lizhili}@umd.edu, xiaowei@pitt.edu, zhe.jiang@ufl.edu, {aolin, shuoxu98}@umd.edu

Societal Importance of Spatial Fairness

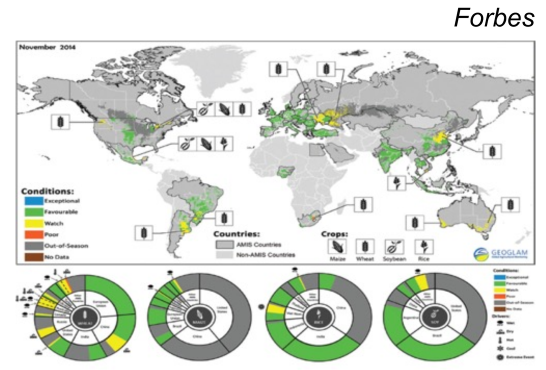
A glimpse of spatial fairness:

	Spatially Fair	Spatially Biased
Overall Accuracy:	80%	80%
Accuracy by location:	80% 80% 80% / 80% 80% 80% / 80% 80% 80%	100% 100% 100% / 100% 40% 100% / 100% 40% 40%

- If left unattended, spatial bias may cause unfair distribution of resources, social division, spatial disparity in sustainability.
- Energy Budget: Temperature estimation
  - Risk of unfair temperature estimation…
  - Risk of unfair heat energy projections…

Forbes

- Agriculture: Crop monitoring
  - Machine learning based crop mapping
    - G20's GEOGLAM, NASA Harvest, …
  - Risk of unfair distribution of subsidies…
- Natural disasters: Flood mapping
  - Risk of unfair distribution of resources…
  - Risk of unfair estimation of insurance rates…
- …

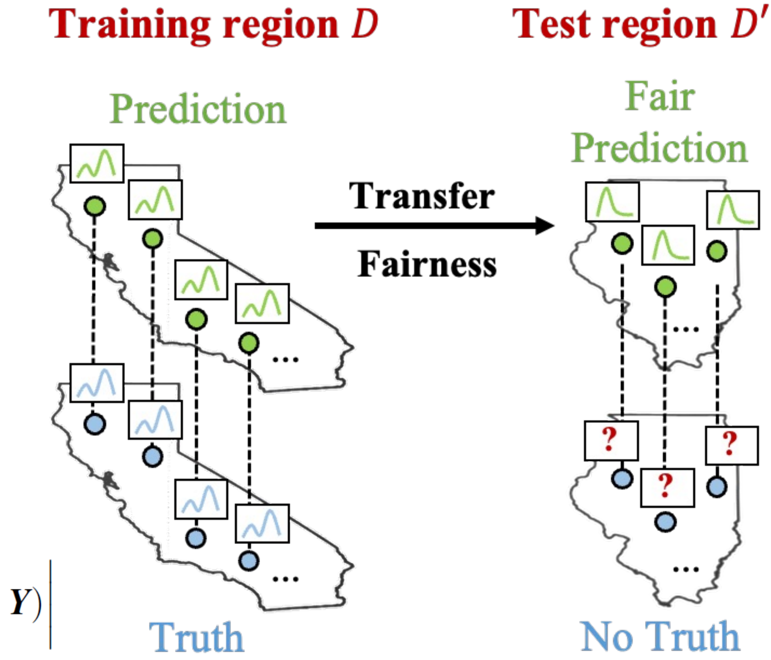


GEOGLAM

Problem Definition

- Inputs
  - X: Input features.
- Outputs
  - Y_{pred}: **fairer** predictions.
- Objectives
  - Prediction performance
    - RMSE, …
  - **Transfer Fairness to test**

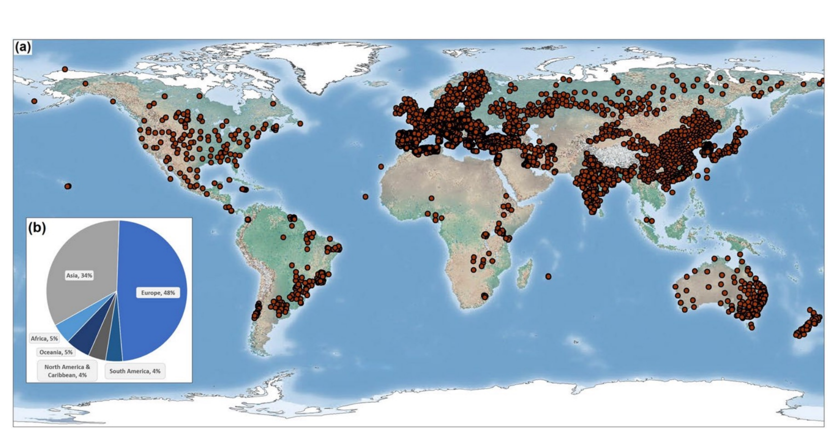
$$\mathcal{L}_f = \frac{1}{|S|}\sum_{s_i \in S}\left|\mathcal{L}_p(\mathcal{F}(x_i), y_i) - \overline{\mathcal{L}_p}(\mathcal{F}(X), Y)\right|$$



Challenges

- Test groups are **not prefixed** and can be **highly dynamic**.
  - Difficult to transfer learning objectives from train to test.
- **No truth** available in test regions.

Distribution of U.S. Climate Reference Network

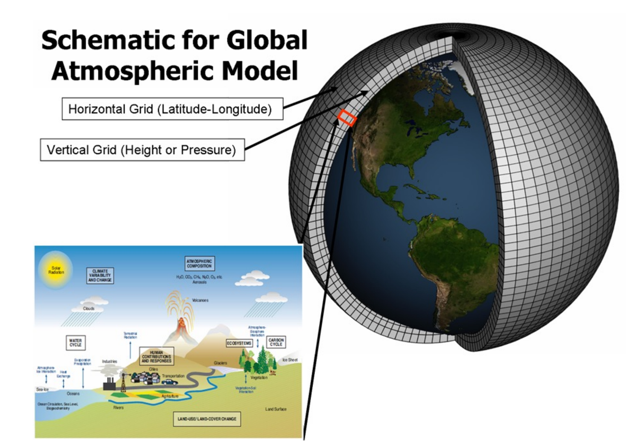


Global distribution of rainfall stations

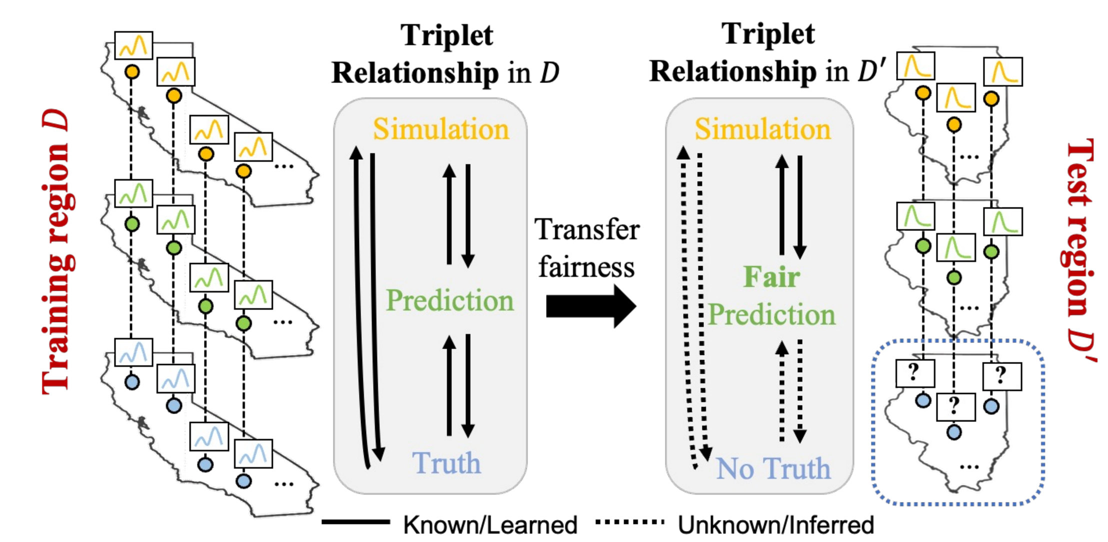
Scientific Report, 2017

Contributions

- A **Dual-fairness consistency** to improve fairness to new locations by reducing the gap between simulations and predictions.
- An **inverse-modeling** design to align the physics-based simulation with the learning objectives for embedding physical laws.
- **Physical-rule-based constraints** to improve prediction performance.

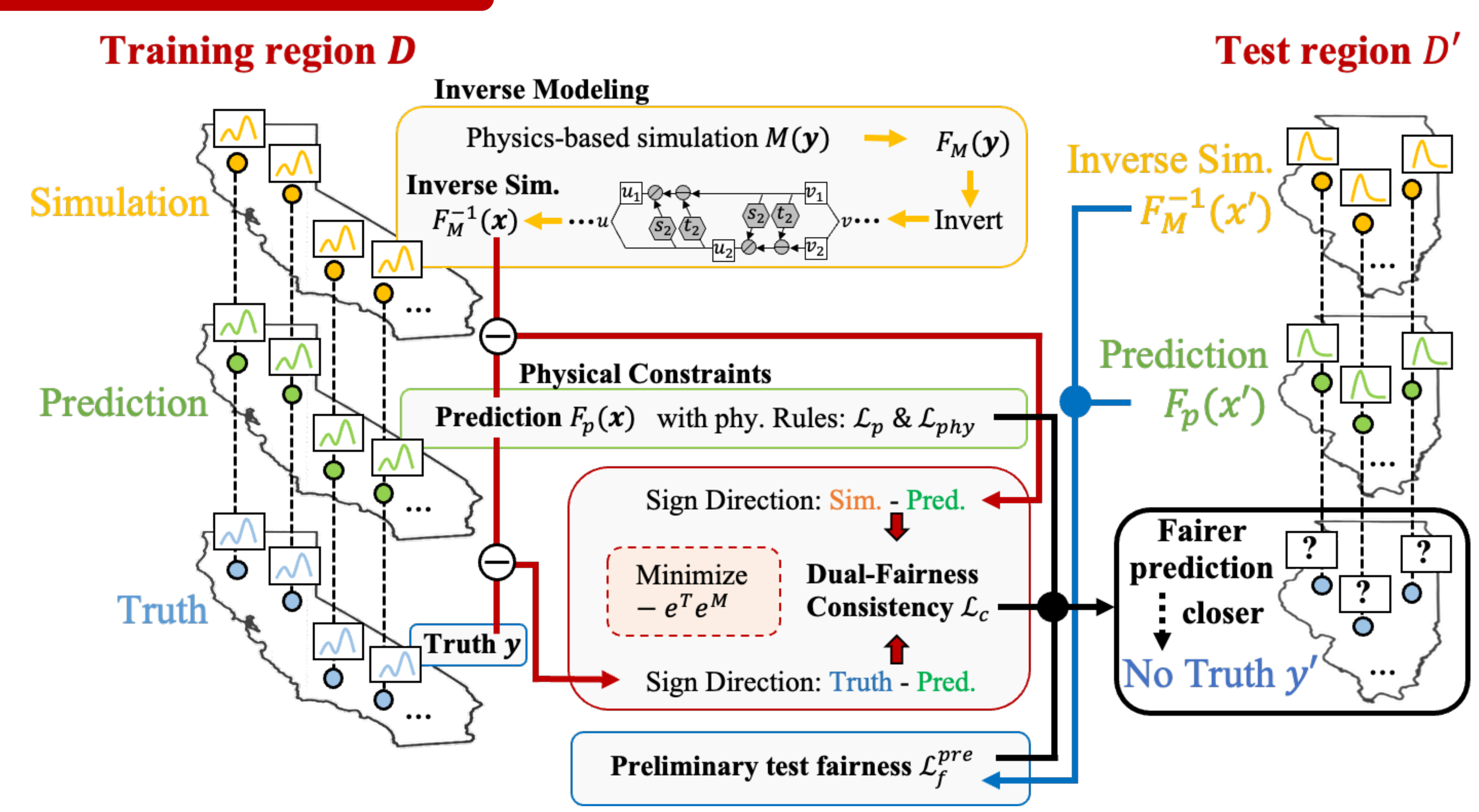


Simulation models: (1) unlimited theoretical labels, (2) flexible spatio-temporal sets of location.



Conceptual idea of fairness relationships.

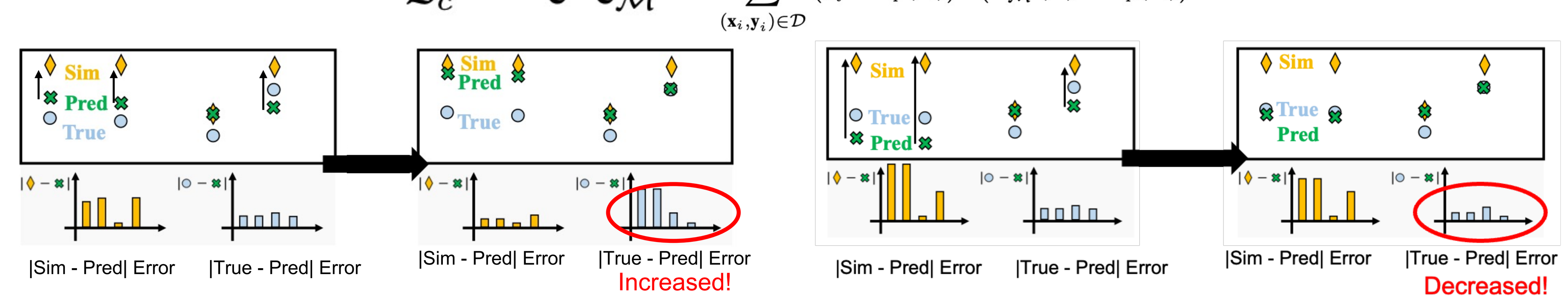
SimFair Framework



Dual-Fairness Consistency

- Goal: to make predictions closer to truth given simulations.
- How: learn **triplet relationships** to have better initial conditions for predictions.

$$\mathcal{L}_c = -\mathbf{e}^T\mathbf{e}_{\mathcal{M}} = -\sum_{(x_i, y_i)\in\mathcal{D}}(y_i - \mathcal{F}_p(x_i)) \cdot (\mathcal{F}_{\mathcal{M}}^{-1}(\mathbf{X}) - \mathcal{F}_p(x_i))$$



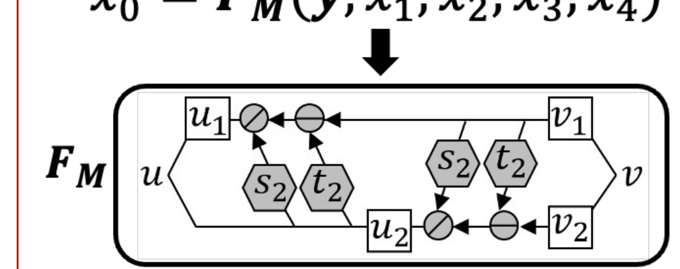
Normal training process. Dual-fairness-guided training process.

Inverse Modeling

A bijector-based invertible network F_M.

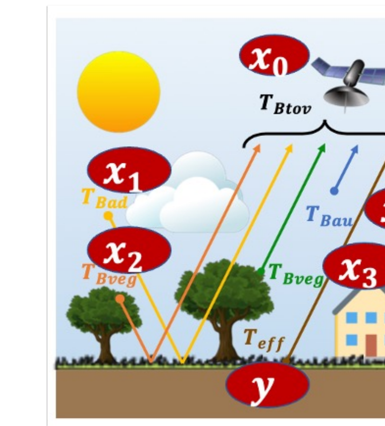
- Forward path: learn **prior physical knowledge** in M.

$$x_0 = F_M(y, x_1, x_2, x_3, x_4)$$



- Inverse path: output target y.

$$y, x_1, x_2, x_3, x_4 = F_M^{-1}(x_0)$$

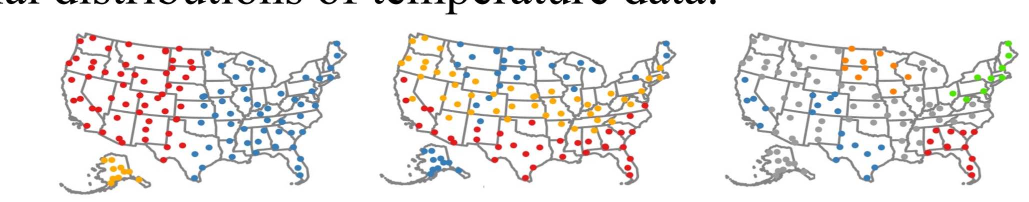


A physical model M.

$$x_0 = M(y, x_1, x_2, x_3, x_4)$$

Experiment Results

Spatial distributions of temperature data:

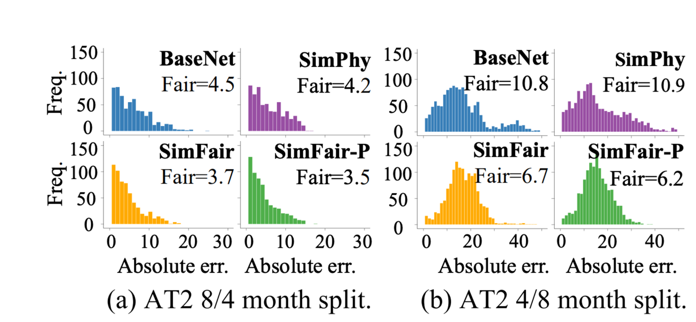


Results:

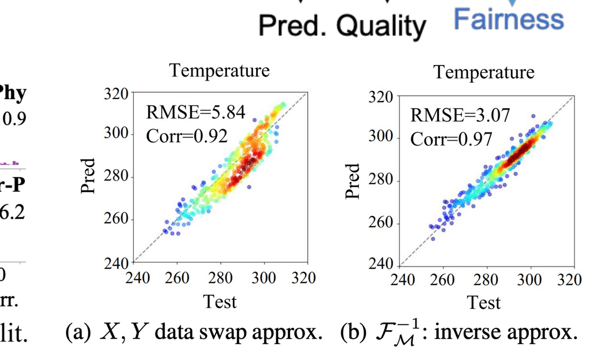
	Model	West-East RMSE	West-East Corr.	West-East Fairness	East-West RMSE	East-West Corr.	East-West Fairness	East-Alaska RMSE	East-Alaska Corr.	East-Alaska Fairness
FNN	BaseNet	6.75	0.83	4.49 (±0.81)	27.56	0.34	13.65 (±2.48)	52.03	0.19	44.42 (±3.37)
	Sim	6.45	0.86	4.69 (±0.82)	20.33	0.44	11.93 (±2.19)	43.74	0.29	38.84 (±5.69)
	SimPhy	7.19	0.84	5.56 (±0.4)	17.78	0.48	10.79 (±1.98)	45.52	0.29	38.84 (±3.4)
	RegFair	7.22	0.8	4.97 (±0.69)	25.35	0.37	12.36 (±2.42)	38.5	0.06	29.73 (±6.78)
	Self-Reg	6.35	0.84	4.27 (±0.7)	31.97	0.31	16.48 (±2.42)	38.01	0.06	28.95 (±4.15)
	SimFair	3.07	0.97	2.04 (±0.19)	3.11	0.96	**1.94 (±0.03)**	6.23	0.84	**4.25 (±0.78)**
	SimFair-P	2.88	0.97	**1.89 (±0.06)**	3.13	0.96	1.96 (±0.05)	6.29	0.81	4.45 (±0.51)
LSTM	BaseNet	4.22	0.93	2.66 (±0.14)	4.02	0.97	2.45 (±0.16)	11.93	0.8	5.14 (±0.31)
	Sim	3.89	0.95	2.43 (±0.15)	3.3	0.97	2.21 (±0.40)	13.32	0.85	5.25 (±0.49)
	SimPhy	4.46	0.95	2.69 (±0.17)	3.23	0.97	2.04 (±0.17)	12.27	0.88	4.82 (±0.28)
	RegFair	4.17	0.94	2.66 (±0.22)	4.03	0.96	2.59 (±0.58)	12.16	0.81	5.03(±0.4)
	Self-Reg	4.10	0.94	2.57 (±0.26)	3.85	0.96	2.41 (±0.16)	11.24	0.84	4.68(±0.41)
	SimFair	3.46	0.96	2.21 (±0.11)	3.22	0.98	**1.91 (±0.11)**	11.05	0.86	4.55(±0.27)
	SimFair-P	3.35	0.96	**2.12(±0.11)**	3.24	0.97	1.99 (±0.17)	10.52	0.89	**4.15(±0.23)**

Proposed Methods

Pred. Quality Fairness



Temporal splits.



Inver modeling.

Conclusions

SimFair: a simulation-guided fairness-aware learning

- transfer fairness to the regions without truth;
- inversely model simulations to align learning objectives;
- softly regularize predictions using physical constraints.

Future: expand to more knowledge- or rule-based models with various domain tasks for fairness-aware learning.

Acknowledgements

NSF Grant No. 2105133 2126474 2147195

NASA Grant No. 80NSSC22K1164 80NSSC21K0314

USGS Grant No. G21AC10207

Google AI for Social Good