ETH zürich MIE lab technische universität dortmund spatial modelling lab dortmund

Estimation of Moran’s I in the Context of Uncertain Mobile Sensor Measurements

Dominik Bucher[1,*], Henry Martin[1,*], David Jonietz[2], Martin Raubal[1], René Westerholt[3,*]
[1]Institute of Cartography and Geoinformation, ETH Zurich, Switzerland
[2]HERE Technologies Switzerland, Zurich, Switzerland
[3]School of Spatial Planning, TU Dortmund University, Germany
*These authors contributed equally to this work
30.09.2021

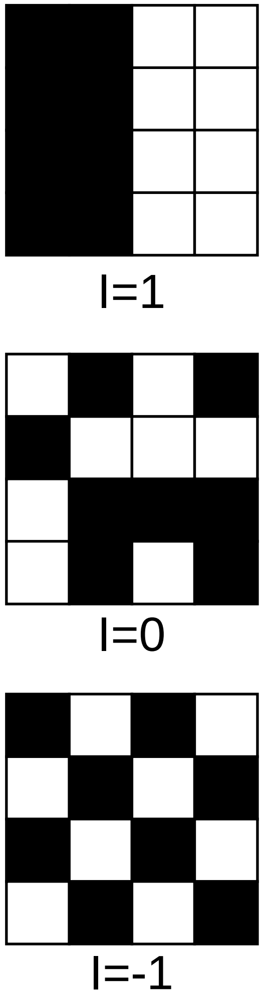
Moran's *I*

$$I = \frac{n}{\sum_{i,j\neq i}^{n} w_{ij}} \cdot \frac{\sum_{i,j\neq i}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i}^{n}(x_i - \bar{x})^2}$$

x_i: observation at location i

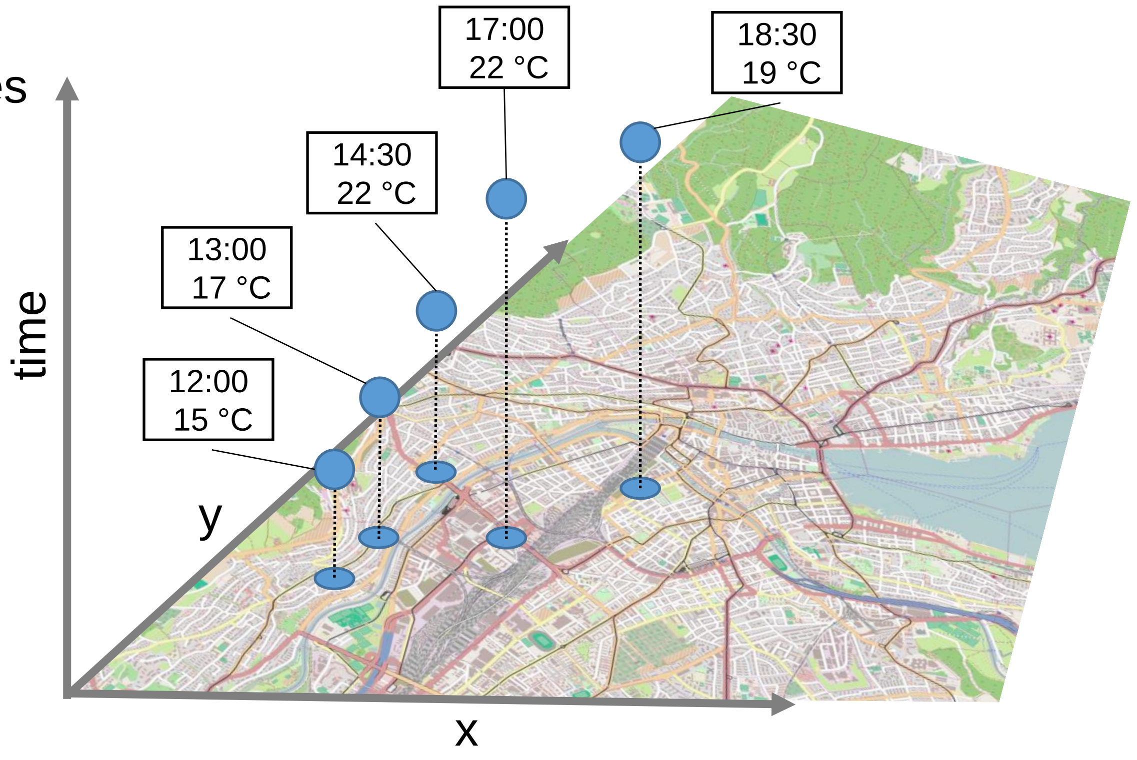
$\bar{x}$: mean of observations

w_{ij}: spatial weight between location i and j



Motivation

- Mobile sensors capture measurements at different locations at different points in time and space.
- Measurements may represent different processes of dynamic geographic phenomena.
- The validity of measurements from the past at different locations is therefore uncertain.
- **Research question:** How can we calculate Moran's *I* with data collected at different times at different locations while taking into account their pair-wise joint uncertainty?



Moran's *I* with Modified Weights

Given a pairwise measure of certainty between the latest observations at two locations u_{ij} we define two sets of new weights which lead to modified Moran's *I* values

$w'_{ij} = w_{ij} \cdot u_{ij}$ → I^1

$w''_{ij} = w_{ij} \cdot (1 + u_{ij} - \bar{u})$ → I^2

Global Moran's *I*

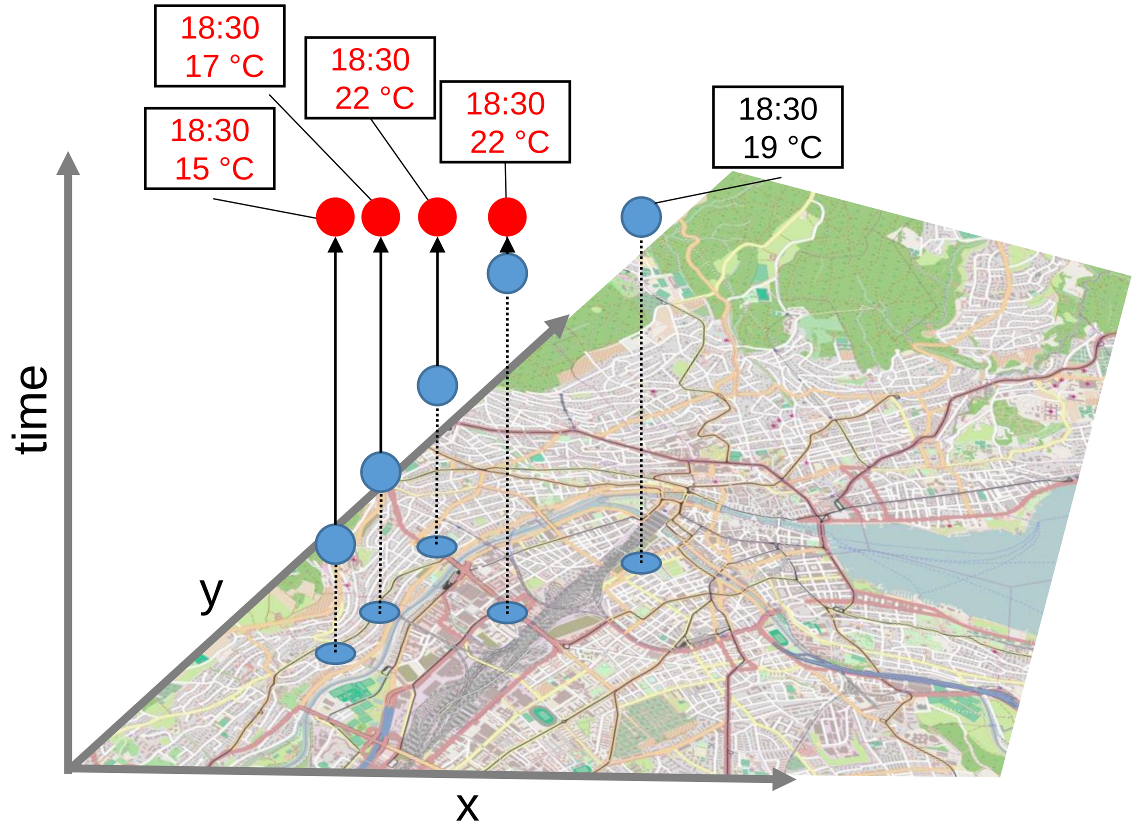
$$I = \frac{n}{\sum_{i,j \neq i}^{n} w_{ij}} \cdot \frac{\sum_{i,j \neq i}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i}^{n}(x_i - \bar{x})^2}$$

Intuition:

- I^1: Absolute approach. Lower the weights of locations with observations with high certainty.
- I^2: Relative approach. Locations with observations with above average certainty have increased weights, those with below average have reduced weights.

Uncertainty Estimation

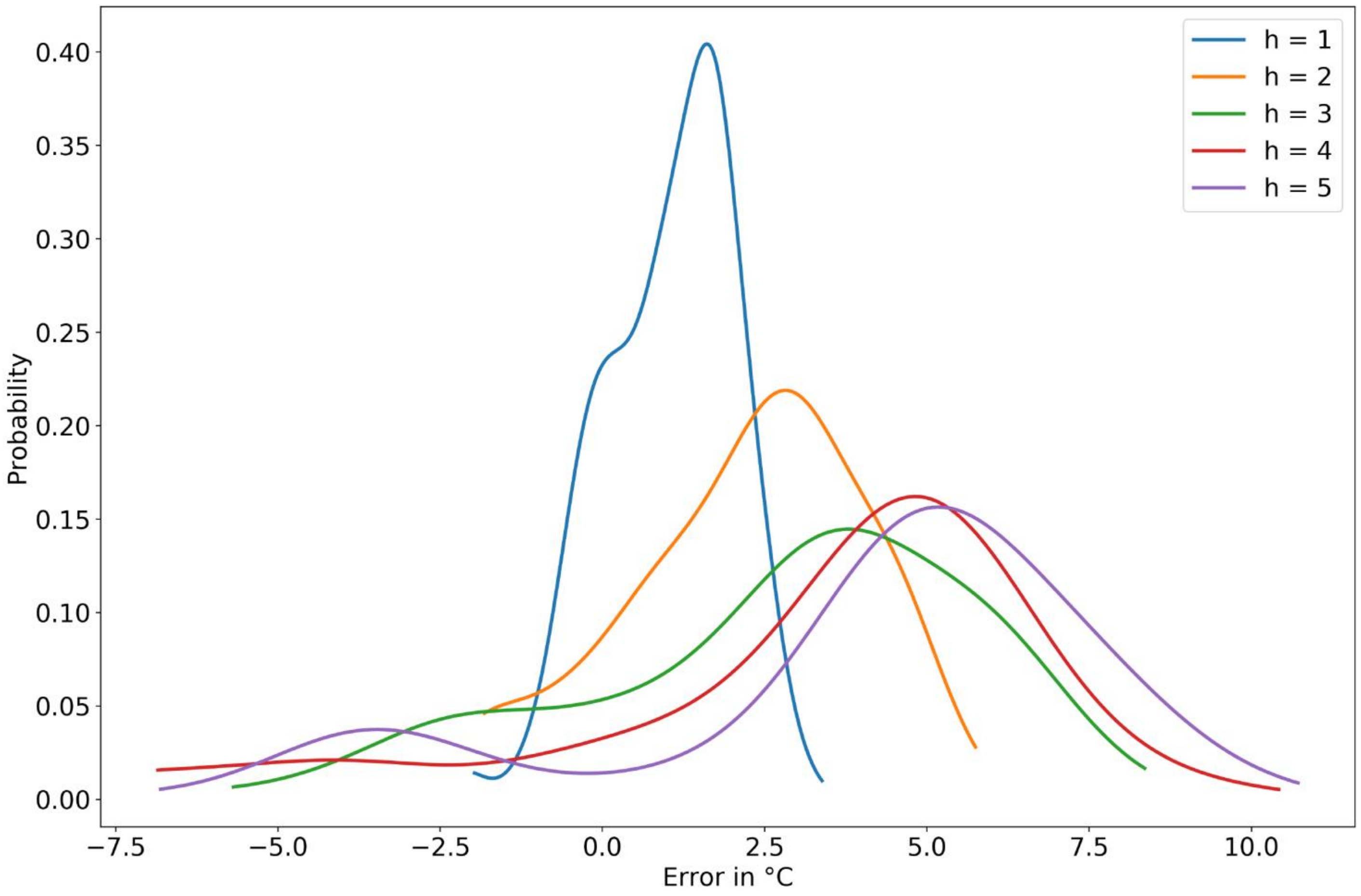
- Create a probabilistic forecast to project all past datapoints to the timestamp of the analysis.
- The probabilistic forecast assigns a distribution to all observations.
- Use shape of distribution to quantify the uncertainty.



Probabilistic Forecast using Empirical Prediction Intervals

- Empirical prediction intervals [1] can be used to transform deterministic forecasts into probabilistic forecasts by adding the distribution of historical forecast errors to the deterministic forecast (dressing).
- Error distributions are location and forecast horizon dependent.

Example of error distributions for different forecast horizons



[1] Yun Shin Lee and Stefan Scholtes. Empirical prediction intervals revisited. *International Journal of Forecasting*, 30(2):217–234, 2014.

Joint Uncertainty Estimation

- Required: Pair-wise **joint (un)certainty** of all observations which can be estimated via the joint distribution of two observations.
- Problem: Via the forecast we only have access to the (non-independent) marginal distributions of each observation. Estimating joint distribution from marginals is often a very complicated problem.
- Hint: Estimating (tight) bounds of the sum of two (potentially) dependent marginal distribution is much easier [1].

$$u_{ij} \leq P(\mathbf{X} + \mathbf{Y} \leq e)$$

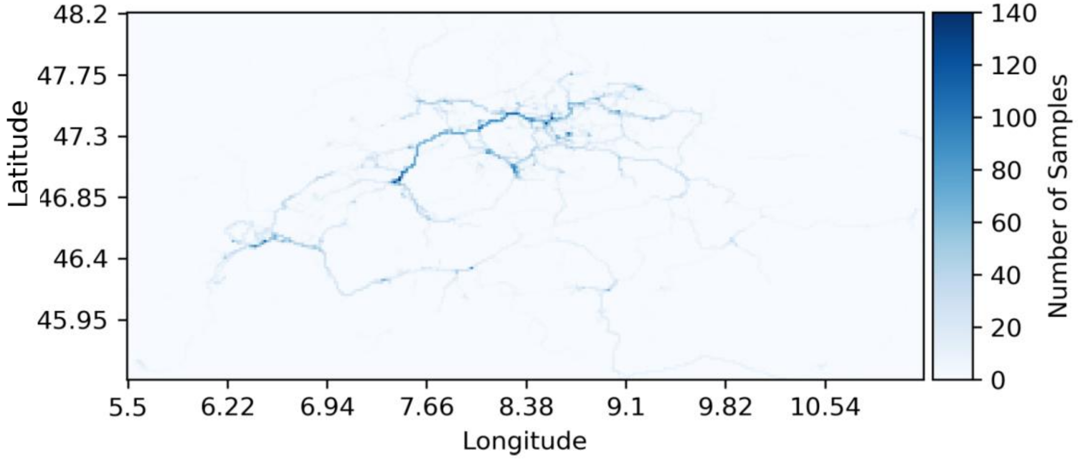
➢ u_{ij} is the (lowest) probability that the sum of the marginals (absolute forecast errors) is below the threshold *e*. Therefore, u is a measurement of joint certainty.

[1] Michel Denuit, Christian Genest, and Étienne Marceau. Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, 25(1):85–104, 1999.

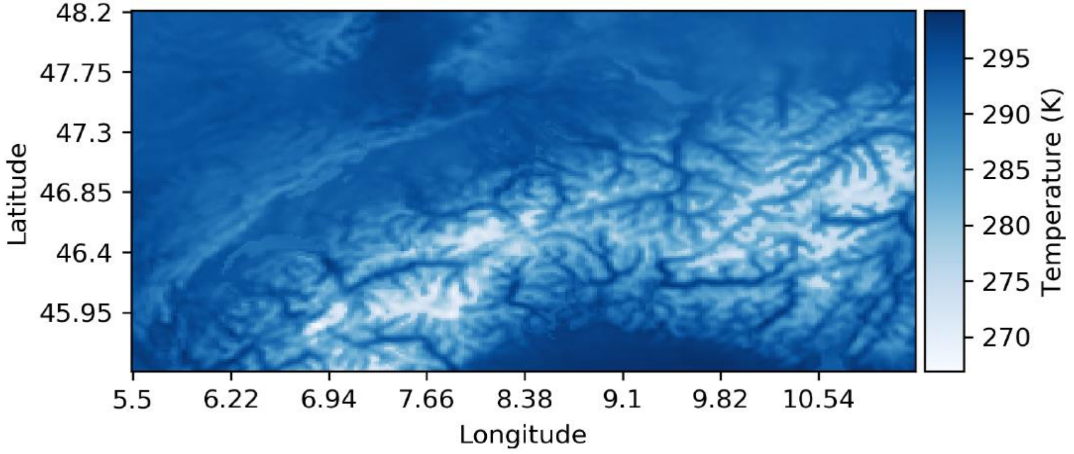
Case Study 1 – Real data: Temperature sensors on cars

- GPS data of car movement from a tracking study in Switzerland are used as moving sensors.
- Car data is used to sample from a 0.018° cell grid (~2x2 km in Switzerland) with ambient temperature data (max 1 sample per cell).

Density of measurements (based on movement data)



Temperature in Switzerland (2nd of July 2013 09:00)



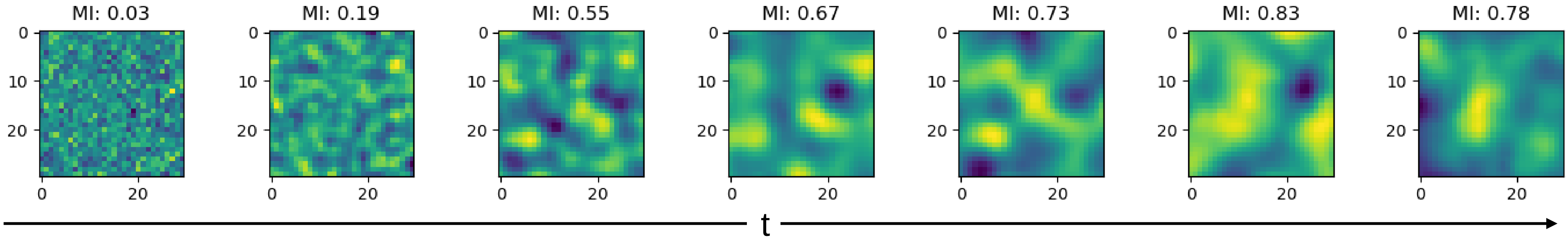
Case Study 2 - Synthetic Data with Spatial Non-Stationarity

- Generate 60 time steps with a 60x60 grid cells resolution.
- Simple Kriging based on a Gaussian spatiotemporal variogram.
- Data has non-stationary spatial autocorrelation.

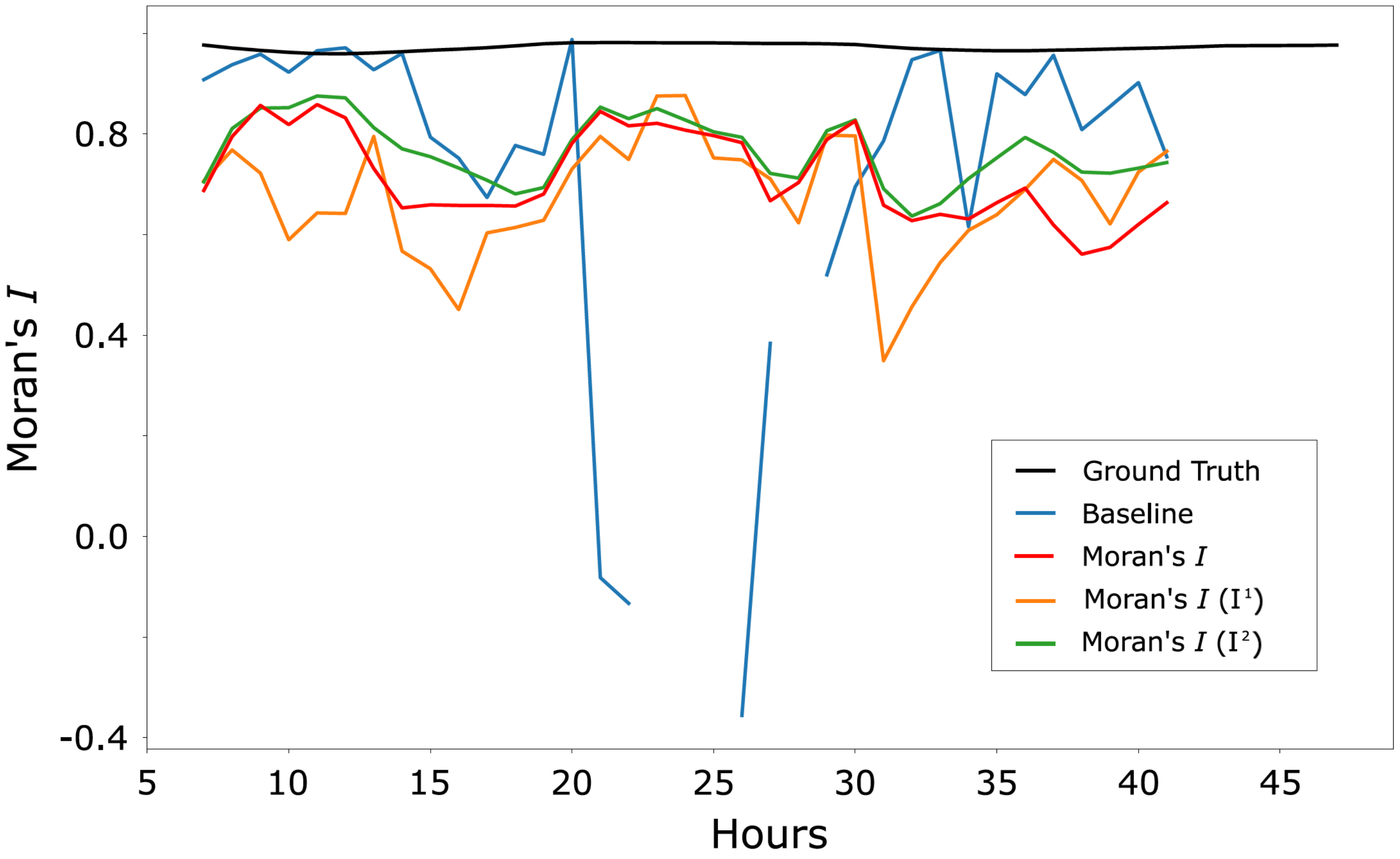
$$\gamma(h, r_i) = (s - n)\left(1 - \exp\left(-\frac{h^2}{\frac{1}{3}r_i^2}\right)\right) + n$$

$$r_i = 0.5 + |10 \cdot \sin(2i/2\pi)|$$

r_i: range
s : sill (1)
n : nugget (0)

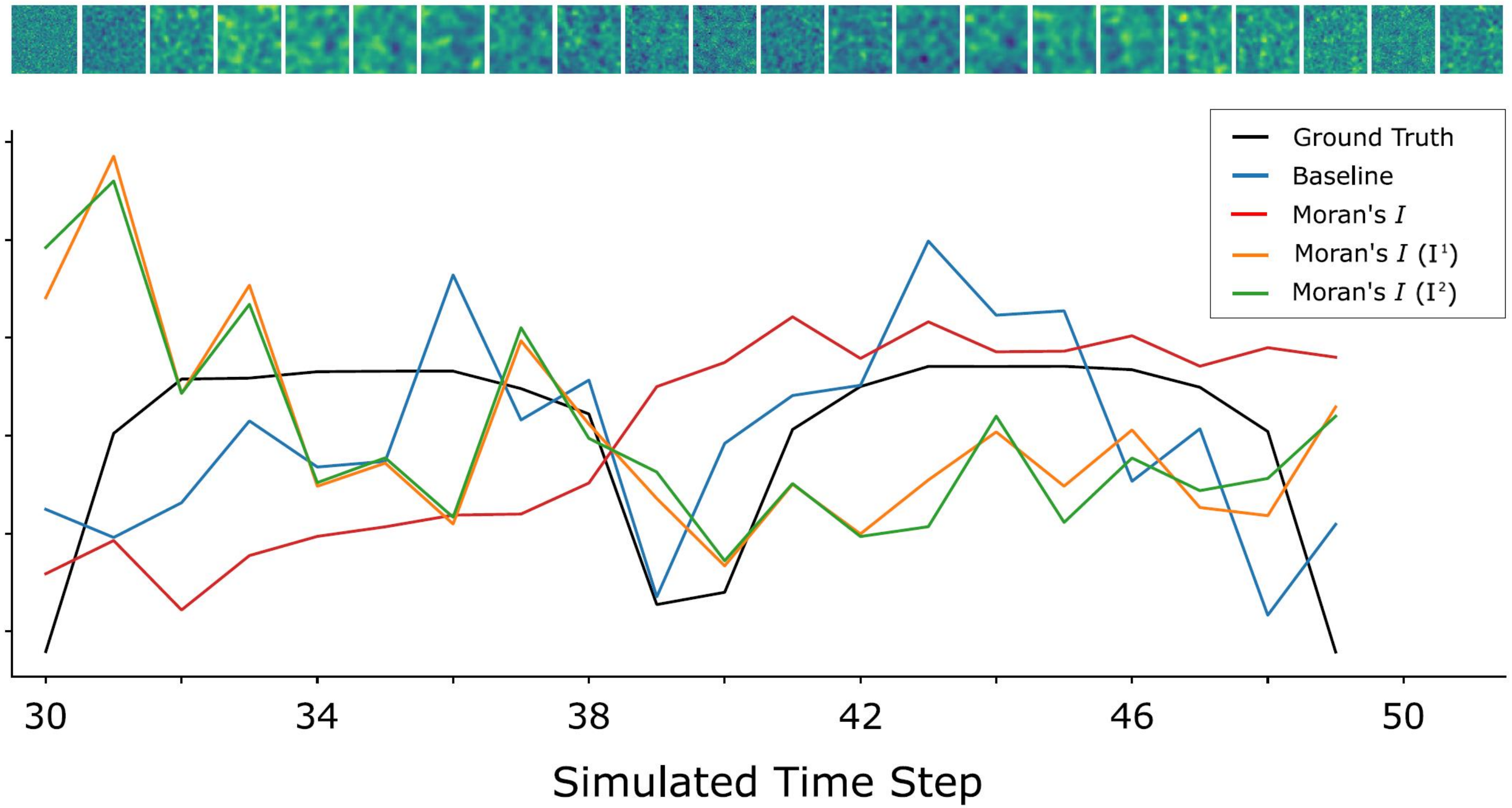


Results Case Study 1



- Moran's I^2 shows improvement over other methods.
- I^1 and I^2 stable in periods with low data availability (causes gaps in baseline).
- As opposed to I^2, I^1 does underestimate Moran's *I* by design. Especially in a regime of low certainty.
- I^2 is less stable and more volatile than I^1.

Results Case Study 2



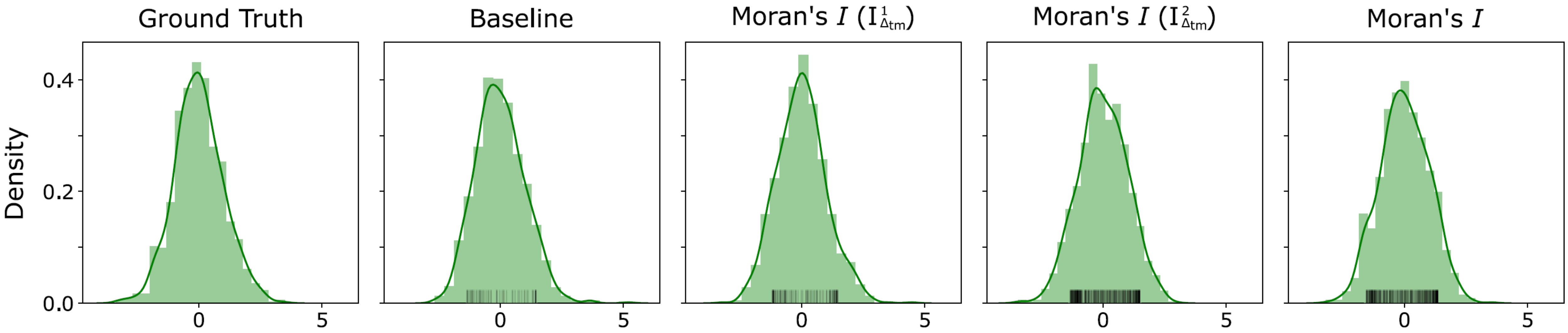
- I^1 and I^2 not suitable for dealing with non-stationarity.
- Due to the low spatio-temporal correlation between time steps, baseline has best performance.

Conclusions & Future Work

- Two novel ways to incorporate uncertainty related to moving sensors in the calculation of Moran's *I*.
- We show that incorporating uncertainty as a relative metric has significant advantages over the approach that incorporates absolute uncertainty values.

- Extend framework to local Moran's *I* to take into account spatial heterogeneity.
- Analyse impact of forecast performance on Moran's *I* estimate.

Back Up


Ground Truth
Baseline
Moran's I ($I^1_{\Delta tm}$)
Moran's I ($I^2_{\Delta tm}$)
Moran's I
Density
0.4
0.2
0.0
0
5

Main Contributions

- Presentation of two modifications of Moran's I that take pair-wise spatio-temporal uncertainty of mobile sensor readings into account.
- Non-parametric approach for the estimation of pair-wise spatio-temporal uncertainty of mobile sensor readings.
- Validation of methods using two case studies.

Probabilistic Forecast using Empirical Prediction Intervals

Empirical prediction intervals [1] can be used to transform deterministic forecasts into probabilistic forecast using historical forecast errors (dressing).

$$\hat{O}_{t_n,h} = f(O_{t \le t_i})$$

$$e_{t_n,h} = O_{t_n} - \hat{O}_{t_n,h}$$

$$\hat{F}_h(e) = k^{-1} \sum_{t=1}^{k} \mathbb{I}(e_{t,h} \le e)$$

$\hat{O}_{t_n,h}$ + $\hat{F}_h(e)$ = probabilistic forecast

[1] Yun Shin Lee and Stefan Scholtes. Empirical prediction intervals revisited. *International Journal of Forecasting*, 30(2):217–234, 2014.

$\hat{O}_{t_n h}$: Forecasted observation for time $t_n$with horizon h
$e_{t_n h}$: Forecast error for time $t_n$with horizon h
$\hat{F}_h(e)$: Empirical cumulative distribution function of Forecast errors

Uncertainty Estimation 2

- Required: Pair-wise **joint uncertainty** of all observations which can be estimated via the joint distribution of two observations.
- Problem: Via the forecast we only have access to the (non-independent) marginal distributions of each observation.
- Approach: Estimate (tight) lower bound of the probability that the forecast error (of the sum of both observations) is below a specified threshold [1].

$$u_{o_{il\to n}, o_{jm\to n}} = b_l(e) = \sup_{x\in\mathbb{R}} \max\{F_1^-(x) + F_2^-(e-x) - 1, 0\} \leq P(\mathbf{X}+\mathbf{Y} \leq e)$$

$u_{o_{il\to n}, o_{jm\to n}}$: Joint certainty of o_{il} and o_{jm} projected to t_n
$b_l(e)$: Lower bound of probability to stay below forecast error
$e_{t_n h}$: Forecast error for time t_n with horizon h

$F_{1/2}^-$: Left side limit of the CDF of the 1st or 2nd marginal distribution

$u_{o_{il\to n}, o_{jm\to n}}$ is the (lowest) probability that the sum of the absolute errors (of both forecasts) is below the threshold *e.* Therefore, *u* is a measurement of certainty.

[1] Michel Denuit, Christian Genest, and Étienne Marceau. Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, 25(1):85–104, 1999.

Uncertainty Estimation 2

- Required: Pair-wise **joint uncertainty** of all observations which can be estimated via the joint distribution of two observations.
- Problem: Via the forecast we only have access to the (non-independent) marginal distributions of each observation.
- Approach: Estimate (tight) lower bound of the probability that the forecast error (of the sum of both observations) is below a specified threshold [1].

$$u_{o_{il \to n}, o_{jm \to n}} = b_l(e) = \sup_{x \in \mathbb{R}} \max\{F_1^-(x) + F_2^-(e - x) - 1, 0\} \leq P(\mathbf{X} + \mathbf{Y} \leq e)$$

$u_{o_{il \to n}, o_{jm \to n}}$: Joint certainty of o_{il} and o_{jm} projected to t_n
$b_l(e)$: Lower bound of probability to stay below forecast error
$e_{t_n h}$: Forecast error for time t_n with horizon h

$F_{1/2}^-$: Left side limit of the CDF of the 1st or 2nd marginal distribution

$u_{o_{il \to n}, o_{jm \to n}}$ is the (lowest) probability that the sum of the absolute errors (of both forecasts) is below the threshold *e*. Therefore, *u* is a measurement of certainty.

[1] Michel Denuit, Christian Genest, and Étienne Marceau. Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, 25(1):85–104, 1999.