SFI Centre for Research Training in Foundations of Data Science

Spatio-temporal Imputation of Missing Rainfall Values to Establish Climate Normals

Brian O'Sullivan
Prof. Gabrielle Kelly

11th May, 2023

HOST INSTITUTIONS

UNIVERSITY OF LIMERICK OLLSCOIL LUIMNIGH

Maynooth University National University of Ireland Maynooth

UCD DUBLIN

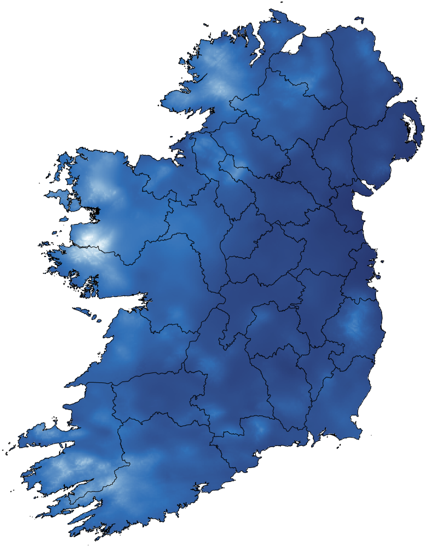
FUNDED BY

Science Foundation Ireland sfi For what's next

Skillnet IRELAND

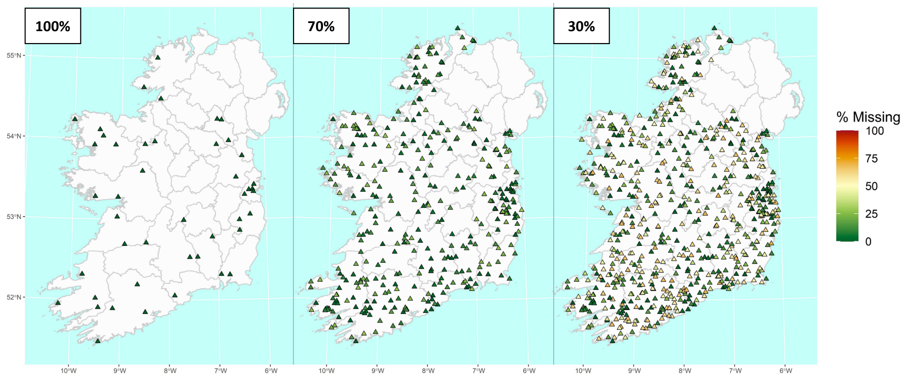
Long-Term Averages (LTAs)

- Republic of Ireland (1981-2010).
- Imputation, Infilling, Gap-filling.
- Complete series during study period is required.



Dataset

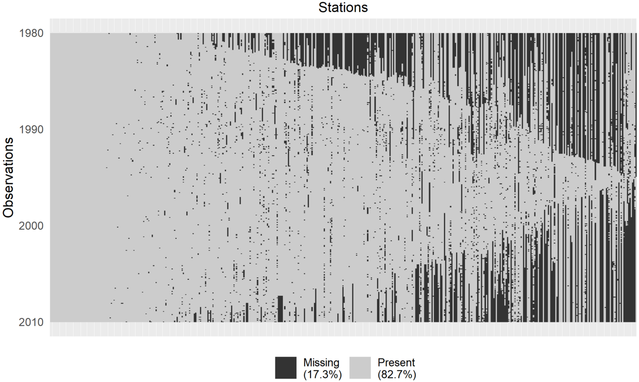
Met Éireann monitoring network



Rain gauge stations (monthly precipitation)

Cutoffs at different station completeness

Data Missingness



High sparsity at the beginning and end of study period

Station Miss	No. Stations	Total Miss	Mean Rainfall	Density (km^2/station)
100%	45	0%	100.30mm	1562.88
70%	365	10.2%	110.13mm	192.68
50%	474	17.3%	108.74mm	148.38
30%	679	30.48%	107.47mm	103.59

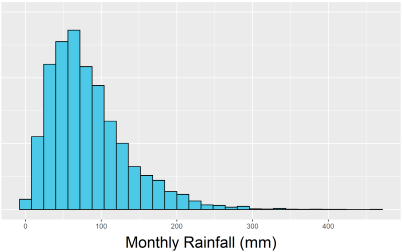
Data Subset

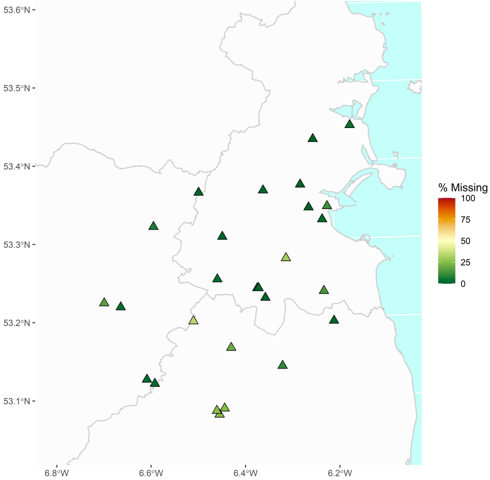
27 Stations - Dublin

70% station completeness

Skewed data

→ square root transformation

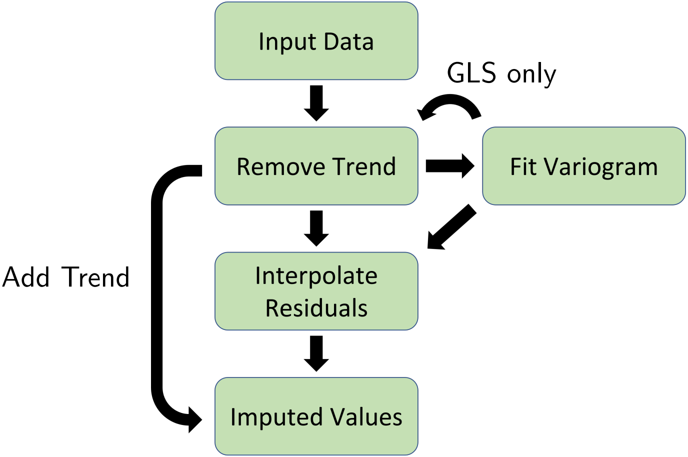




Method Breakdown

Kriging assumes a Gaussian spatio-temporal random field Z

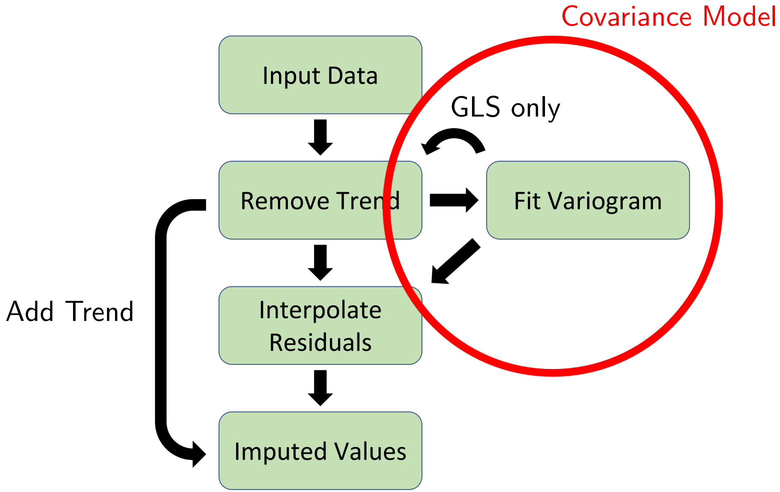
First remove any trends $\rightarrow$ Regression



Method Breakdown

Kriging assumes a Gaussian spatio-temporal random field Z

First remove any trends $\rightarrow$ Regression



Regression - Regularisation

Linear Model: $Y = \beta_0 + X\beta + \epsilon$

Covariates X: east, north, elev, time (+ second-order interactions)

Ordinary Least Squares:

$$\min_{\beta_0,\, \beta} \left\{ \|Y - \beta_0 - X\beta\|^2 \right\}$$

Add Regularisation Penalty, L:

$$\min_{\beta_0,\, \beta} \left\{ \tfrac{1}{N} \|Y - \beta_0 - X\beta\|^2 + L \right\}$$

Lasso Penalty: $L_1 = \lambda\|\beta\|_1$ Ridge Penalty: $L_2 = \lambda\|\beta\|^2$

Regression - Elastic-Net Regularisation

Elastic-Net Penalty: $L_{ENet} = \lambda \left(\alpha \|\beta\|_1 + \|(1-\alpha)\beta\|^2 \right)$

$$\min_{\beta_0,\, \beta} \left\{ \|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2 + \lambda \sum_{j=1}^{p} (\alpha \hat{\beta}_j^2 + (1-\alpha)|\hat{\beta}_j|) \right\}$$

$\lambda \rightarrow$ causes shrinkage of regression coefficients β

$\alpha \rightarrow$ spans Ridge ($\alpha = 1$) and Lasso ($\alpha = 0$) regression

Enables variable selection (Lasso)

Shrinks β of highly correlated covariates (Ridge)

Regression - Generalised Least Squares

Ordinary Least Squares (OLS) assumes independent residuals, ϵ

Generalised Least Squares (GLS) accounts for dependency between ϵ

OLS: $\epsilon \sim N(0, \sigma^2 I)$ GLS: $\epsilon \sim N(0, C)$

OLS Estimator: $\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$

GLS Estimator: $\hat{\beta} = \left(X^T C^{-1} X\right)^{-1} X^T C^{-1} Y$

Covariance matrix, C, modelled from variogram:

$$2\gamma(s_i, s_j) = E[(Z(s_i) - Z(s_j))^2]$$

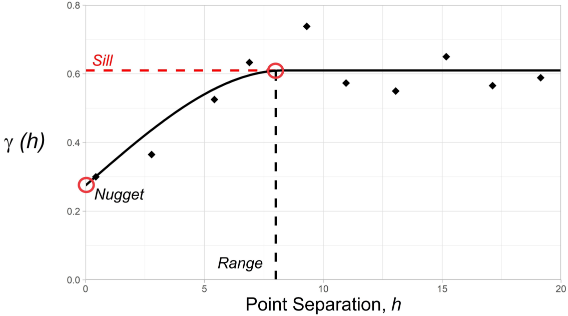
Variogram - Spatial

Assume Z is isotropic & stationary, $\rightarrow C(s_i, s_j) = C(\|s_i - s_j\|) = C(h)$

i.e., correlation between point pairs only depends on separation

Autocorrelation:

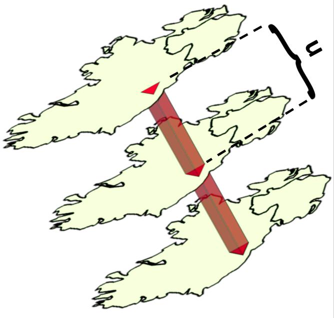
$$\rho(h) = C(h)/C(0) \qquad \gamma(h) = \tau^2 + \sigma^2(1 - \rho(h))$$



Variogram - Spatio-Temporal

Consider temporal separation between points, u

Consider dataset across entire study period



Spatio-temporal metric (anisotropy $\rightarrow \kappa$)

Separable covariance model: $C_{sep}(h, u) = C_s(h)C_t(u)$

Sum-metric covariance model:

$$C_{sm}(h, u) = C_s(h) + C_t(u) + C_{joint}(\sqrt{h^2 + (\kappa \cdot u)^2})$$

Variogram - Spatio-Temporal

Matérn correlation structure ρ (extra shape parameter ν):

$$\rho(h) = \{2^{\kappa-1}\Gamma(\nu)\}^{-1}(h/\phi)^{\nu}K_{\nu}(h/\phi)$$

Sum-metric model $\rightarrow$ spatial, temporal, and joint Matérn structures
Thirteen total parameters:

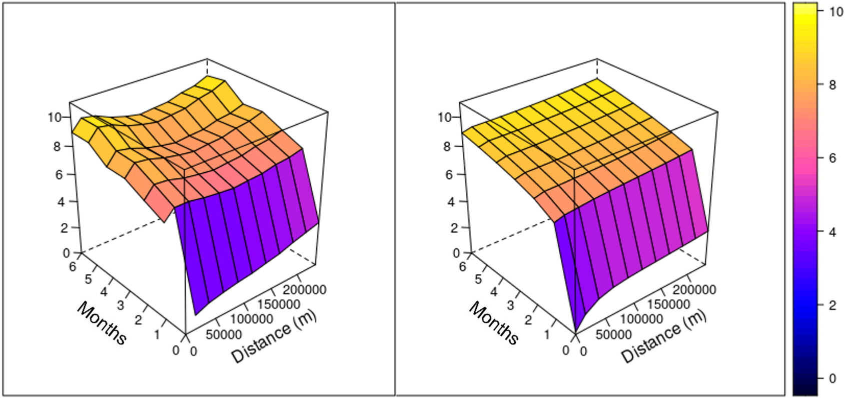
sill$\times$3, nugget$\times$3, range$\times$3, shape$\times$3, anisotropy

Surface fitting implemented with gstat and L-BFGS-B algorithm

Variogram - Spatio-Temporal

Sum-metric Spatio-Temporal Variogram

$$\gamma(h,u) = \gamma_s(h) + \gamma_t(u) + \gamma_{joint}\left(\sqrt{h^2 + (\kappa \cdot u)^2}\right)$$



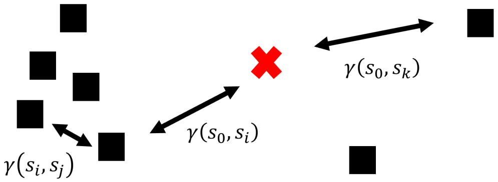
Kriging Interpolation

Imputation → Predicted values from weighted average:

$$\hat{Y}(s_0) = \sum_{i=1}^{N} w_i \; Y(s_i)$$

Ordinary kriging system:

$$\gamma(s_i, s_j)w = \gamma(s_0, s_i) \qquad \sum_{i=1}^{N} w_i = 1$$



Validation Tests

10-Fold Cross Validation:

- Inverse Distance Weighting
- Spatial Regression-Kriging
- Spatio-temporal Regression-Kriging
- Elastic-Net Spatio-temporal Regression-Kriging
- GLS Spatio-temporal Regression-Kriging (Only on 27 stations)

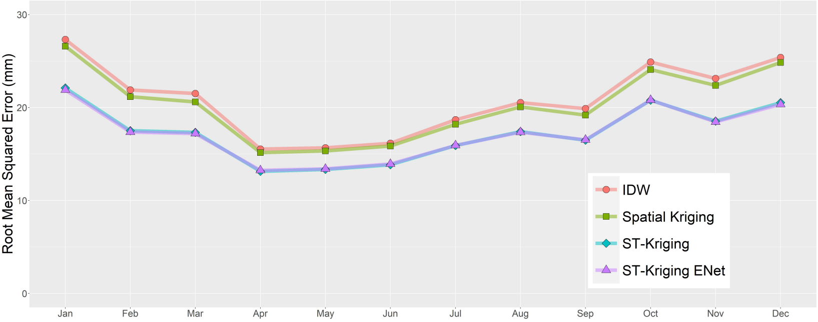
$$w_{idw}(s_0) = \frac{1}{\|s_0 - s_i\|^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \quad ; \quad RMSE_r = \frac{RMSE}{\sigma_{obs}}$$

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

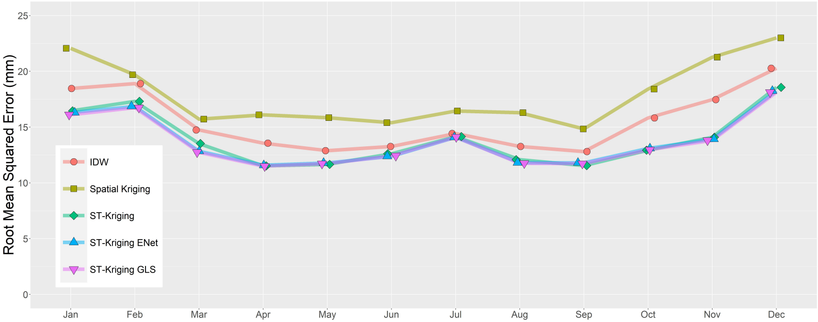
Results - 474 Stations (50% completeness cutoff)

Method	RMSE	$RMSE_r$	R^2
IDW	21.21	30.28	0.909
Spatial Kriging	20.61	29.41	0.914
ST-Kriging	17.47	24.94	0.938
ST-Kriging ENet	17.42	24.86	0.938



Results - 27 Stations (Dublin)

Method	RMSE	$RMSE_r$	R^2
IDW	15.69	29.81	0.911
Spatial Kriging	18.13	34.44	0.884
ST-Kriging	14.05	26.70	0.929
ST-Kriging ENet	13.89	26.39	0.930
ST-Kriging GLS	13.80	26.23	0.931

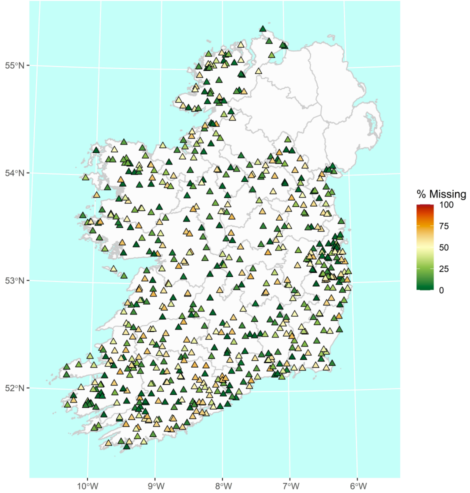


Completeness Cutoffs

Imputation via
Elastic-Net ST-Kriging

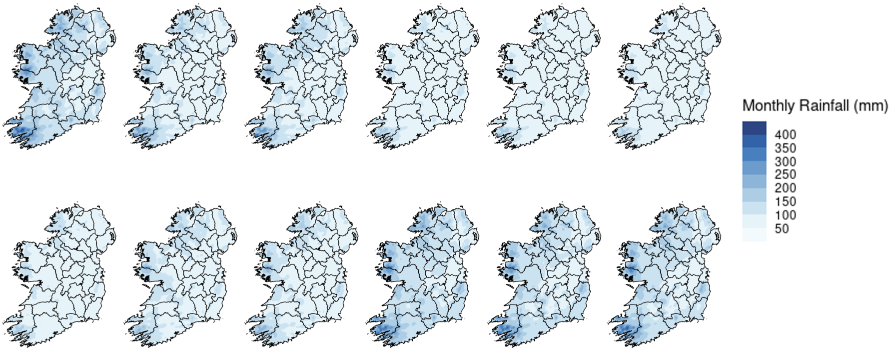
% Missing	RMSE	$RMSE_r$	R^2
70%	17.89	25.18	0.937
50%	17.42	24.86	0.938
30%	17.15	24.86	0.938

Future exploration of optimal completeness cutoff



Monthly LTAs

Final monthly averages interpolated on 1km grid
(Elastic-Net Spatial Kriging)



Conclusions

- Imputation allows for denser network when producing climate normals.
- Spatio-temporal vs. spatial methods evaluated for precipitation.
- Regularisation offers simple improvement for little additional cost.
- Intractability of GLS for large datasets (Inversion of large C).

Many thanks:

- Prof. Gabrielle Kelly, UCD.
- Met Éireann Climate Services and Research Division, Ireland.

Thank You!

References

Gräler, B., Pebesma, E., and Heuvelink, G. (2016). "Spatio-Temporal Interpolation using gstat". The R Journal 8 (1), 204–218.

Friedman, J. H., Hastie, T., and Tibshirani, R. (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent". Journal of Statistical Software 33 (1), 1–22.

Walsh, S. (2016). Long-term rainfall averages for Ireland, 1981-2010. 15, Climatological Note. Met Éireann.

HOST INSTITUTIONS

UNIVERSITY OF LIMERICK OLLSCOIL LUIMNIGH

Maynooth University National University of Ireland Maynooth

UCD DUBLIN

FUNDED BY

Science Foundation Ireland sfi For what's next

Skillnet IRELAND