

**Spatio-temporal prediction of missing temperature with stochastic Poisson equations  
The LC2019 team winning entry for the EVA 2019 data competition**

Cheng, Dan; Liu, Zishun

**DOI**

[10.1007/s10687-020-00397-w](https://doi.org/10.1007/s10687-020-00397-w)

**Publication date**

2020

**Document Version**

Final published version

**Published in**

Extremes

**Citation (APA)**

Cheng, D., & Liu, Z. (2020). Spatio-temporal prediction of missing temperature with stochastic Poisson equations: The LC2019 team winning entry for the EVA 2019 data competition. *Extremes*, 24(1), 163-175. <https://doi.org/10.1007/s10687-020-00397-w>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.



# Spatio-temporal prediction of missing temperature with stochastic Poisson equations

The LC2019 team winning entry for the EVA 2019 data competition

Dan Cheng<sup>1</sup> · Zishun Liu<sup>1</sup>

Received: 31 March 2020 / Revised: 13 October 2020 / Accepted: 20 October 2020 /  
Published online: 17 November 2020  
© The Author(s) 2020

## Abstract

This paper presents our winning entry for the EVA 2019 data competition, the aim of which is to predict Red Sea surface temperature extremes over space and time. To achieve this, we used a stochastic partial differential equation (*Poisson equation*) based method, improved through a regularization to penalize large magnitudes of solutions. This approach is shown to be successful according to the competition's evaluation criterion, *i.e.* a threshold-weighted continuous ranked probability score. Our stochastic Poisson equation and its boundary conditions resolve the data's non-stationarity naturally and effectively. Meanwhile, our numerical method is computationally efficient at dealing with the data's high dimensionality, without any parameter estimation. It demonstrates the usefulness of stochastic differential equations on spatio-temporal predictions, including the extremes of the process.

**Keywords** Data competition · EVA 2019 Conference · Prediction · Poisson equation · Spatio-temporal data · Temperature data

**AMS 2000 Subject Classifications** 62P12 · 62H11 · 62M30 · 35Q62

## 1 Introduction

The aim of EVA 2019 data competition is to predict spatio-temporal extremes of Red Sea surface temperature (Huser 2020). The dataset of the competition consists of daily sea surface temperature (SST) anomalies for the entire Red Sea from 1985 to

---

This work was first presented on July 5th, 2019, at EVA 2019.

✉ Zishun Liu  
z.liu-4@tudelft.nl

<sup>1</sup> Delft University of Technology, Delft, The Netherlands

2015 whilst part of the data is masked. The main goal of this data competition is to predict the distribution of

$$X(s, t) = \min_{(\tilde{s}, \tilde{t}) \in \mathcal{N}(s, t)} \hat{A}(\tilde{s}, \tilde{t}), \quad (1)$$

for space-time validation points  $(s, t)$ , where  $\hat{A}(\cdot, \cdot)$  is the predicted SST anomalies. The competition evaluates the performance via a threshold-weighted continuous ranked probability score (twCRPS). The main challenge is the data's high dimensionality (16703 locations and 11315 days) and strong non-stationarity in space and time (Section 3.3 of Huser (2020)). More details of the data competition, including the definition of the anomaly, plots of the data along with some exploratory analysis, as well as the competition's rules and goals, are available in the Editorial (Huser 2020).

To predict the extremes on the spatio-temporal domain, it is sufficient to predict the whole field of values on it. That is to say, we can predict  $\hat{A}(\cdot, \cdot)$  instead of directly predicting  $X(\cdot, \cdot)$  in Eq. (1). Given the partial data of  $\hat{A}(\cdot, t)$  for someday  $t$ , the predicted data in the unobserved region should comply with some continuity constraints. According to Fourier's analytical theory of heat, it is reasonable to assume a temperature field has a second order derivative, implying that the anomaly field has its Laplacian. We formulate the missing value prediction problem as a Poisson equation based model. The Poisson equations are defined on the two-dimensional spatial domain, not the spatio-temporal domain. For each time step (day), Poisson equations are solved to interpolate the unobserved regions. To generate the stochasticity in predictions, our Poisson equation is not deterministic but stochastic. The stochastic components (*i.e.* the Laplace fields) in our SPDEs (stochastic partial differential equations) are sampled from the observed data, assuming temporal stationarity, which allows bypassing explicit estimation of parameters defining a stochastic model. The method is further improved with a regularization term to penalize large magnitudes of solutions. Experimental results show that it is effective and efficient to use the regularized stochastic Poisson equation to fill in the missing data in this problem.

A lot of approaches have been proposed on spatio-temporal extremes in the field of extreme value theory (EVT). One category of approaches is to model block maxima using max-stable processes (Davison et al. 2012; Davis et al. 2013; Huser and Davison 2014). The other category is to model the exceedances of a threshold (Pickands 1975; Bacro et al. 2020) using extreme distributions such as generalized Pareto distribution (GPD). Both categories of models have been used on climate data (Chavez-Demoulin and Davison 2005; Reich and Shaby 2012). SPDE-based approaches have been made popular by the INLA method in spatial and spatio-temporal statistics (Krainski et al. 2018). In Lindgren et al. (2011), a type of SPDE similar to our Poisson equations is studied, which will be discussed in the last section of this paper. Comparing to EVT-based methods, our method is more efficient since we do not need any parameter estimation. The drawback is its performance will be inferior comparing to methods tailored to the prediction of the upper tail. Our approach is very similar to the Poisson image editing method (Pérez et al. 2003) from the image processing community, which blends multiple images seamlessly with Poisson equations.

The remainder of this paper is structured as follows. In Section 2, we give the details about the stochastic Poisson equation based model. Its numerical method and the regularized version are given in Section 3. In Section 4, we explain our scheme to generate predictions, and the cross-validation dataset we prepared to evaluate our approach, followed by some concluding discussions in Section 5.

**Notation** The following notation is used: *Italic lower-case letters* refer to scalar values ( $x$ ) or scalar-valued functions ( $f(\cdot)$ ). *Lower-case bold letters* denote vectors ( $\mathbf{x}$ ) or vector-valued functions ( $\mathbf{f}(\cdot)$ ). Finally, *upper-case bold letters* represent matrices ( $\mathbf{X}$ ).

## 2 Method

The missing data prediction problem can be studied from the perspective of data interpolation. By considering a 1D curve  $y = f(x)$  with missing interval  $(x_0, x_1)$ , as shown in Fig. 1, the interpolation result should have at least  $C^0$  continuity. It is straightforward to connect two endpoints of the missing interval with a straight line segment, on which the second derivative is zero. Thus finding such a line segment is equivalent to solving the differential equation

$$\begin{cases} \frac{d^2 f(x)}{dx^2} = 0, & x \in (x_0, x_1) \\ f(x_0) = y_0, & f(x_1) = y_1. \end{cases} \tag{2}$$

The one-dimensional case above can be easily generalized to 2D, *i.e.* assuming the second derivative on the missing region is zero. Given a function  $u(x, y)$ ,  $(x, y) \in \mathcal{S}$  with masked region  $\mathcal{M} \subset \mathcal{S}$ , the unobserved region can be interpolated by solving the equation

$$\begin{cases} \Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, & (x, y) \in \mathcal{M} \\ u(x, y) \text{ takes known value in } \mathcal{S} \setminus \mathcal{M}. \end{cases} \tag{3}$$

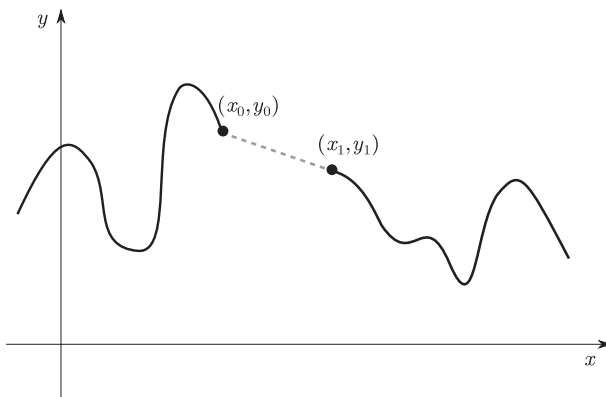


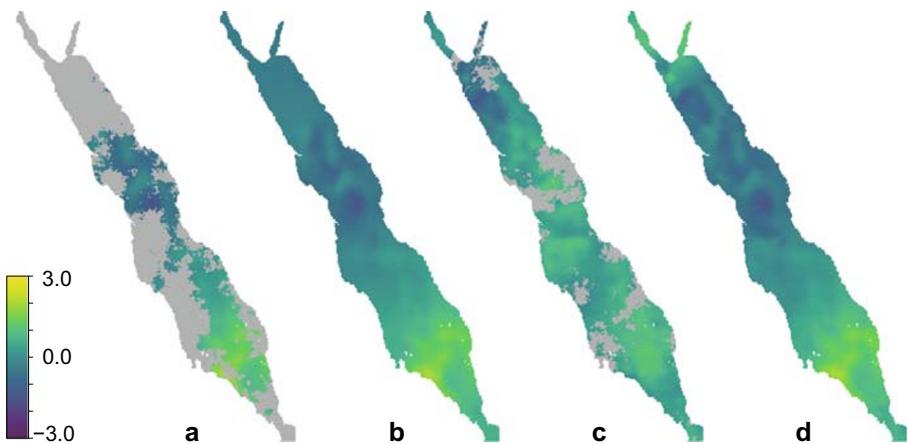
Fig. 1 One-dimensional data interpolation

Equations taking the form  $\Delta u = 0$  are known as Laplace equations. An interpolation example is demonstrated in Fig. 2 (a–b). It is observed that the predicted region is over-smooth, lacking variations, compared to the known area, indicating that the right-hand side (RHS) in Eq. (3) should be some function  $f$  which is not zero. Thus the missing region can be predicted by solving the equation  $\Delta u = f$ , which is known as Poisson equation.

The physical meaning of Poisson equation  $\Delta u = f$  here is that, given the partial temperature data  $u$  of someday  $t \in \mathcal{T}$ , the Laplacian of  $u$  (we call it Laplace field) in the unobserved region should follow some target field  $f$ . Immediately, it is necessary to model the target Laplace field  $f$  conditioning on the known data. We assume  $f$  is random since its exact value is unknown. Then the equation we are solving is actually a stochastic Poisson equation, instead of a deterministic one. Although we can construct some statistical model for random  $f$ , it is more straightforward to sample existing fields from known data directly, with the assumption that the Laplace field is independent and identically distributed (i.i.d.) in time for the sake of simplicity. That is to say, the Laplace fields of day  $t'_1, t'_2, \dots$  are independent and identically distributed realizations of the random function  $f$ . The stochastic Poisson equation is solved by solving a series of deterministic ones. In terms of solving one of the Poisson equations, we take a reference day  $t' \in \mathcal{T}$  as an example. Day  $t'$  is picked as a reference day, then Laplace field of day  $t'$  is a reference target Laplace field for day  $t$ . Therefore, missing data interpolation for day  $t$  can be accomplished by solving

$$\begin{cases} \Delta u(x, y; t) = \Delta u(x, y; t'), & (x, y) \in \mathcal{M}_t \\ u(x, y; t) = \text{known, in } \mathcal{S} \setminus \mathcal{M}_t, \end{cases} \quad (4)$$

in which  $\mathcal{M}_t$  is the unobserved region of the day  $t$ . In general, this function does not always have a unique solution, but our numerical method gives a unique solution in our specific setting, which is described in Section 3. There are still missing data in



**Fig. 2** Given the partial data of Dec 22, 2015 **a**, the unobserved regions (in gray) are interpolated with Laplace equation, which is over-smooth **b**. With the data of Apr 1, 1985 as reference **c**, the solution of Poisson equation is more realistic **d**

$\Delta u(x, y; t')$  since  $u(x, y; t')$  is incomplete for any  $t'$ . These values in  $\Delta u(x, y; t')$  will be simply filled with *zeros*. An example is shown in Fig. 2 (c–d).

As indicated in Huser (2020), the temperature anomalies are not stationary in space and time. The Poisson equation  $\Delta u = f$  can handle it effectively. In fact, the solution of Poisson equation can be decomposed as the sum of the solutions of the following two equations,

$$\begin{cases} \Delta u = 0, & (x, y) \in \mathcal{M} \\ u(x, y) = \text{known, in } \mathcal{S} \setminus \mathcal{M}, \end{cases} \quad \text{and} \quad \begin{cases} \Delta u = f, & (x, y) \in \mathcal{M} \\ u(x, y) = 0, \text{ in } \mathcal{S} \setminus \mathcal{M}. \end{cases} \quad (5)$$

In these two equations, the left one is the Laplace equation that imposes the continuity constraints, and the right one provides the randomness. The left one, which is only related to the boundary conditions, ensures the solution is adaptive to the known data of the current day, tackling the non-stationarity in *time*. The right one has RHS  $f = \Delta u(x, y; t')$ , making the Laplace field be the same as some other day in a location-wise sense. Thus it resolves the temperature anomalies' non-stationarity in *space*. In conclusion, the sum of them can handle the non-stationarity in both space and time.

### 3 Numerical solution

#### 3.1 Numerical solution of the Poisson equation

The numerical solution of Poisson equation on a rectangular domain with Dirichlet boundary condition can be found with the finite difference method (Knabner and Angerman 2003). The key idea is discretizing the domain as a grid and approximating the Laplace operator with a linear combination of neighbor grid nodes. The differential equation is then converted to a linear system, of which the left-hand side (LHS) matrix is named the Laplacian matrix. In this paper, Poisson equation Eq. (4) is solved with the finite difference method.

First, the spatial domain is represented as a mesh grid  $\Omega$ . The Red Sea  $\mathcal{S}$  has been discretized into  $S = 16703$  grid cells in the dataset of this competition. The grid cells are represented with their row/column indices  $(i, j)$ , in which  $i \in \{1, \dots, 359\}$ ,  $j \in \{1, \dots, 233\}$ . Each cell  $(i, j)$  has at most four neighbors  $(i \pm 1, j)$  and  $(i, j \pm 1)$ . Then the mesh grid  $\Omega$  is established after taking cells  $(i, j)$  as mesh nodes and the neighboring relation as mesh edges.

Then, the Laplacian matrix is determined after defining the mesh grid  $\Omega$ . Noting that the mesh grid  $\Omega$  is equivalent with an undirected graph  $\mathcal{G}$  with its nodes and edges, the Laplacian matrix  $\mathbf{L}$  of graph  $\mathcal{G}$  is defined according to Godsil and Royle (2001), as a generalization of the Laplacian matrix of a rectangular domain. The Poisson equation Eq. (4) is converted to the linear system

$$\mathbf{L}\mathbf{u} = \mathbf{L}\mathbf{u}', \quad (6)$$

in which matrix  $\mathbf{L}$  has shape  $S \times S$ ,  $\mathbf{u}$  is an  $S$ -dimensional column vector containing the temperature anomalies of all locations on the day  $t$ , and  $\mathbf{u}'$  is the temperature anomalies vector of some other day  $t'$ . In this equation, the LHS  $\mathbf{L}\mathbf{u}$  is the Laplace

field of day  $t$ , and the RHS  $\mathbf{L}\mathbf{u}'$  is the reference Laplace field taken from day  $t'$ . Both of them are computed with the Laplacian matrix  $\mathbf{L}$  defined on the mesh domain. As discussed above, some elements in  $(\mathbf{L}\mathbf{u}')$  are not available since  $\mathbf{u}'$  is incomplete (containing values represented with NaNs). These invalid values in  $(\mathbf{L}\mathbf{u}')$  are filled with zeros.

Finally, unknown values of  $\mathbf{u}$  are determined by solving the equation. Without loss of generality, it can be assumed that vector  $\mathbf{u}$  is split into two parts  $\mathbf{u}_0$  and  $\mathbf{u}_1$ , of which the first one  $\mathbf{u}_0$  contains all unobserved locations while the known values appear in the second part  $\mathbf{u}_1$ . This form can be obtained by multiplying permutation matrices on both sides of Eq. (6). Then  $\mathbf{L}$  and  $\mathbf{u}'$  are partitioned accordingly, resulting in the linear system

$$\begin{pmatrix} \mathbf{L}_{00} & \mathbf{L}_{01} \\ \mathbf{L}_{10} & \mathbf{L}_{11} \end{pmatrix} \begin{pmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \end{pmatrix} = \begin{pmatrix} \mathbf{L}_{00} & \mathbf{L}_{01} \\ \mathbf{L}_{10} & \mathbf{L}_{11} \end{pmatrix} \begin{pmatrix} \mathbf{u}'_0 \\ \mathbf{u}'_1 \end{pmatrix}, \quad (7)$$

which implies that

$$\mathbf{L}_{00}\mathbf{u}_0 = \mathbf{L}_{00}\mathbf{u}'_0 + \mathbf{L}_{01}(\mathbf{u}'_1 - \mathbf{u}_1). \quad (8)$$

This equation has a unique solution since  $\mathbf{L}_{00}$  is invertible, which will be proved in the next subsection. At last, Eq. (8) can be solved with LU decomposition of matrix  $\mathbf{L}_{00}$ .

### 3.2 An optimization perspective: least squares

From the viewpoint of mathematical optimization, the Poisson equation Eq. (4)

$$\Delta u = \nabla \cdot \nabla u = \nabla \cdot \nabla u' \quad (9)$$

is the Euler-Lagrange equation of the type

$$\min_u \iint \|\nabla u - \nabla u'\|^2. \quad (10)$$

A discrete approximation of this optimization problem is

$$\min_{\mathbf{u}} \|\mathbf{G}\mathbf{u} - \mathbf{G}\mathbf{u}'\|^2, \quad (11)$$

in which  $\mathbf{G}$  is the gradient matrix. The gradient matrix is defined as transpose of graph  $\mathcal{G}$ 's incidence matrix, when an arbitrary *orientation* is given to  $\mathcal{G}$  (Godsil and Royle 2001). Rows and columns of matrix  $\mathbf{G}$  are the numbers of edges and nodes in graph  $\mathcal{G}$  respectively. If  $\mathbf{G}$  is written in the block form according to values' availability as before, *i.e.*  $\mathbf{G} = (\mathbf{G}_0 \ \mathbf{G}_1)$ , finding unknown values  $\mathbf{u}_0$  is to solve the least squares problem

$$\min_{\mathbf{u}_0} \|\mathbf{G}_0\mathbf{u}_0 - \mathbf{h}\|^2, \quad (12)$$

in which  $\mathbf{h} = \mathbf{G}_0\mathbf{u}'_0 - \mathbf{G}_1(\mathbf{u}'_1 - \mathbf{u}_1)$ . Invalid values (NaNs) in  $\mathbf{h}$  are also filled with zeros.

The least squares problem (12) is solved with the pseudo inverse theory, *i.e.*

$$\mathbf{u}_0 = (\mathbf{G}_0^\top \mathbf{G}_0)^{-1} \mathbf{G}_0^\top \mathbf{h}. \quad (13)$$

Noting that  $\mathbf{L}_{00} = \mathbf{G}_0^\top \mathbf{G}_0$  since  $\mathbf{L} = \mathbf{G}^\top \mathbf{G}$ , solution of the least squares problem should be identical with that of Eq. (8) if there is no missing value on the reference

day. When there is missing data, the only difference between them is that it is the invalid gradient values that are filled with zeros in Eq. (13), while it is the invalid Laplacian values that are replaced with zeros in Eq. (8).

To make  $L_{00} = G_0^T G_0$  invertible, block  $G_0$  in the gradient matrix is required to have linearly independent columns. This is true because each unknown node on graph  $\mathcal{G}$  is connected with at least one known node by a path (Godsil and Royle 2001), which can be concluded from the fact that the spatial domain  $\Omega$  has only one connected component and all days have at least one known node.

### 3.3 Least squares with regularization

Solution of the least squares problem in Eq. (12) gives a prediction of the missing data that is smooth and vibrant (Fig. 3 (c)). However, it is suffering from numerical issues. Although the Laplacian matrix  $L_{00}$  is invertible, its condition number is large from the viewpoint of computing, making the numerical solution unstable. If the Poisson equation has a closed domain with Dirichlet boundary condition, the solution will be well bounded by known values. However, this is not our case.

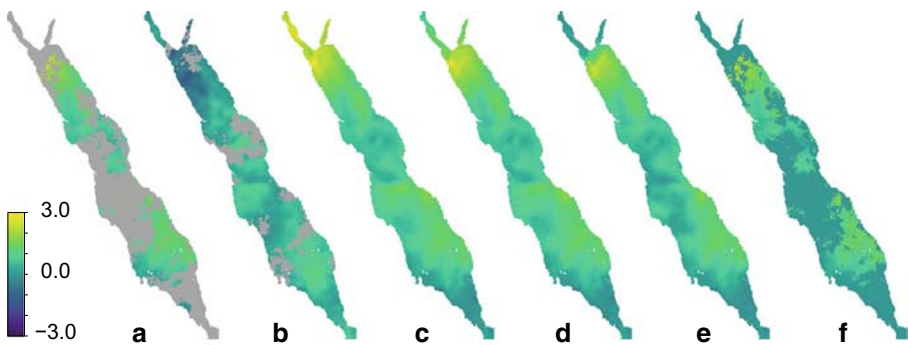
Regularization is commonly used to stabilize least squares, based on which Eq. (12) becomes

$$\min_{\mathbf{u}_0} \|\mathbf{G}_0 \mathbf{u}_0 - \mathbf{h}\|^2 + \lambda \|\mathbf{u}_0\|^2, \tag{14}$$

with  $\lambda \geq 0$ . The regularized least squares problem takes the form of ridge regression (Hastie et al. 2009), of which the solution is given by

$$(\mathbf{L}_{00} + \lambda \mathbf{I}) \mathbf{u}_0 = \mathbf{f}. \tag{15}$$

Here,  $\mathbf{f}$  is  $G_0^T \mathbf{h}$ , and  $\mathbf{I}$  is the identity matrix. Please note that  $\mathbf{f}$  can also take the value of RHS in Eq. (8), depending on how invalid values at the RHS are handled, as discussed in the last subsection. When  $\mathbf{f}$  takes the RHS of Poisson equation, Eq. (15)



**Fig. 3** Results of the regularized least squares. Given the partial data of Sept 23, 2015 **a** and the reference data of Apr 1, 1985 **b**, the spatial domain is completed with the regularized least squares method, using  $\lambda = 0.0, 0.001, 0.01, +\infty$  respectively **c-f**



actually solves  $(\Delta + \lambda)u = f$ , which is named *screened Poisson equation*. It is the Euler-Lagrange equation of the type

$$\min_u \iint \|\nabla u - \nabla u'\|^2 + \lambda \|u\|^2. \quad (16)$$

This minimization objective means the gradient field of the target day should be close to that of the reference day. At the same time, large values of  $u$  are penalized.

It is necessary to determine the value of  $\lambda$  in the regularized problem. Observing the results with different  $\lambda$  values in Fig. 3 (c–f), we can see that larger  $\lambda$  generates flatter result. By considering the limit situation when  $\lambda \rightarrow +\infty$ , the predicted temperature anomalies will degenerate to all zeros (Fig. 3 (f)). In this case, there is a step between observed and unobserved regions (“boundary step”), which can be evaluated as follows

$$\text{BS}(\lambda; t) = \sum_{(p,q) \in P} \|\mathbf{u}_0(p) - \mathbf{u}_1(q)\|^2, \quad (17)$$

where  $P$  is the set of all neighboring location pairs  $(p, q)$  where  $p$  is unobserved and  $q$  is observed. Unavailable data  $\mathbf{u}_0$  is uniquely determined once given  $\lambda \geq 0$ , thus  $\mathbf{u}_0$  is a function of  $\lambda$ . So is the “boundary step”  $\text{BS}(\lambda; t)$ . The “boundary step” is large when the regularization parameter  $\lambda$  is too large or too small. However, finding the optimal  $\lambda$  is not trivial. We pick the value of  $\lambda$  that minimizes the “boundary step” among all candidates in a finite set  $\Lambda$ , *i.e.*

$$\lambda^*(t) = \arg \min_{\lambda \in \Lambda} \text{BS}(\lambda; t). \quad (18)$$

The candidate set  $\Lambda$  will be given numerically in the next section. Please note that the best weight  $\lambda^*$  is determined for each day  $t$  separately. We do not estimate a fixed value of  $\lambda$  for all days.

## 4 Inference

To predict the distribution of  $X(s, t)$  (see Eq. (1)) for space-time validation point  $(s, t)$ , we will predict the whole temperature anomaly field  $\hat{A}(s, t)$  for all days that fall in the  $\pm 3$  days temporal neighborhoods of validations days first, then compute the spatio-temporal neighborhood minimums directly. And last, we use the empirical distribution of the minimums as the prediction.

### 4.1 Inference procedure

Given a space-time validation point  $(s_0, t_0)$ , the temperature anomaly field of that target day,  $\hat{A}(s, t_0)$ ,  $s \in \mathcal{S}$ , is predicted first. To achieve this, a guidance day  $t'$  is picked to provide a reference of Laplace or gradient field. The sampled day  $t'$  should be as complete as possible, thus it is sampled from the period 1985–2006 with only 20% missing data. After that, linear systems are constructed and solved following the method presented in the previous section. All space-time validation points from the day  $t_0$  are computed together from the same linear system, to avoid redundant computing, which saves time, energy, and helps fight climate change.

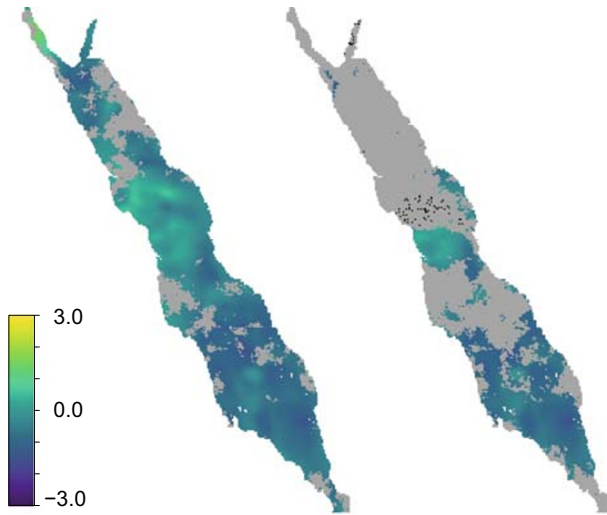
After predicting the field of the target day  $t_0$ , we repeat the same procedure to predict all seven days in its temporal neighborhood, *i.e.* the days  $t_0 + i$ ,  $-3 \leq i \leq 3$ . Hence seven reference days should be sampled. Instead of choosing them independently, it is better to sample seven successive days, *i.e.* the days  $t' + i$ ,  $-3 \leq i \leq 3$ . The advantage is that the temporal dependency of the temperature anomalies is implicitly maintained in the data of sampled seven days. Until now, all  $\hat{A}(\cdot, \cdot)$  values in  $\mathcal{N}(s_0, t_0)$  are obtained. Then we calculate the minimum to get one realization of  $X(s, t)$ .

An empirical distribution of  $X(s, t)$  is predicted if the whole procedure above is performed repeatedly for  $t'_1, t'_2, \dots, t'_N$ . We repeated  $N$  times to generate  $N$  values of  $X(s, t)$  for each validation point  $(s, t)$ , and used the empirical cumulative distribution function (ECDF) of them as the prediction of that validation point. It is noted that the LHS Laplacian matrix in Eq. (6–8) only depends on the data availability, *i.e.* the binary mask, of the target day; its temperature anomaly values and that of the reference day only affect the RHS of the linear systems. We take advantage of this to pre-factorize the Laplacian matrix with LU decomposition only once. Then all of those days sharing the same data availability can be solved directly by forward and backward substitutions.

## 4.2 Leave-one-out validation

Since the validation set of this competition is not available to the teams, a cross-validation set is prepared during the competition. Data in the range of 1985–2006 is used for 22-fold cross-validation. For the  $i$ -th fold ( $i = 0, \dots, 21$ ), data of the year  $1985 + i$  is left out for validation while the other 21 years are used for training. For each day in the left-out year, some of the locations have already been unobserved in the original training set (gray regions in the left of Fig. 4) and more locations are masked out (gray regions in the right of Fig. 4) for cross-validation. For example, as demonstrated in Fig. 4, 1985 is left out and part of Jan 5's data is masked out (new gray regions in the right figure comparing to the left). Unmasked regions (colored regions in the right figure) are used as training data in the cross-validation stage. For some points in the gray region, they can be sampled as cross-validation points if their spatio-temporal neighborhood  $\mathcal{N}(\cdot, \cdot)$  is completely available in the original training set, which are shown as black dots in the right of Fig. 4. Ground truth  $X(\cdot, \cdot)$  is available on these points. Therefore, we have generated the training set and validation points for cross-validation. Since the average temperature anomalies of the first 22 years are lower than that of the last 9 years (where the competition's final predictions should be made), as indicated in Huser (2020), we translated the data by the differences of empirical means between the last 9 years and our cross-validation set in order to match the means.

We evaluated different configurations of our method on the generated cross-validation set, including Poisson equation Eq. (8), regularized least squares Eq. (15), and screened Poisson equation that takes Eq. (15)'s LHS and Eq. (8)'s RHS. For those configurations with regularizers, the candidate set of weight  $\lambda$  is  $\Lambda = \{0.02 \times (0.5)^k : k = 0, 1, \dots, 11\}$ . For each day, we choose the  $\lambda$  value that gives the least “boundary step” as Eq. (18). Consequently, the Poisson equation generates one solution for



**Fig. 4** Cross-validation set is constructed by masking part of the training data. Left: Training data of Apr 5, 1985. Right: Gray regions are masked. Colored regions are the training data for cross-validation. Black dots are cross-validation points

each day; each of the regularized methods, *i.e.* the regularized least squares and the screened Poisson equation, chooses one solution from  $|\Lambda| = 12$  candidate solutions for each day. At last, an ensemble is proposed, which pools all 25 resulting candidate  $\mathbf{u}_0$ s (one from the Poisson equation and 12 from each of the regularized methods) and picks the one with the least “boundary step” from them as Eq. (18). In all methods above, the sample size  $N$  is 1000. These methods are compared to the benchmark proposed in Huser (2020), as shown in Tab. 1. The ensemble gives the lowest error in terms of the average twCRPS on the cross-validation set, thus we used its result as the final submission for the competition.

After sampling  $N = 1000$  times, the ECDF curves are already smooth enough. We have tried to fit the ECDFs with the generalized extreme value (GEV) distribution, while there is little difference according to the twCRPS.

**Table 1** Average twCRPS of different methods on the cross-validation set

Method	$10^4 \times \overline{\text{twCRPS}}$
Benchmark	5.37
Poisson equation	2.46
Regularized least squares	2.33
Screened Poisson equation	2.35
Ensemble	2.19

### 4.3 Final submission

Our approach is implemented with Python to perform parallel computing with an Intel® Core™ i7-9700K CPU on a desktop computer. In our final submission to the competition, we used the ensemble method with sample size  $N = 1000$  and candidate weight set  $\Lambda$  as described above. The total computational cost is less than 70 minutes.

## 5 Discussion

This approach is shown to be successful in the context of this competition, thanks to its advantages. First, the Laplacian matrix is a nice encoding of the spatial domain's geometry. It contains the links between locations and its neighbors, and helps to build the spatial dependence between locations. Second, the temporal dependence is considered implicitly via borrowing the Laplace fields from successive days, *i.e.* reference weeks, instead of independent reference days. Third, the non-stationarity over time and space is embedded in our model since the values in unobserved regions can be decomposed into two parts (Eq. (5)); one is associated only with the current day's information and the other considers the spatial variation. Last, this method is training-free. Training data are used directly as references without any time-consuming parameter estimation.

In Lindgren et al. (2011), the SPDE  $(\lambda - \Delta)^{\alpha/2}u = f$  is studied, where  $\alpha > 0$ ,  $\lambda \geq 0$  and  $f$  corresponds to Gaussian white noise. First, it should be noted that the  $-\Delta$  in Lindgren et al. (2011) is actually the  $+\Delta$  in this paper. The latter notation is used here to make it compatible with the discretization of Laplace operator in Godsil and Royle (2001). Therefore, the above SPDE coincides with the (screened) Poisson equation in this work, when  $\alpha = 2$ ,  $\lambda \geq 0$ . The difference is that our model is derived from an interpolation perspective. Our regularization term with  $\lambda$  is motivated by penalizing large magnitudes of solutions. At the same time, our model is parameter-free since the value of  $\lambda$  is determined according to the "boundary step" on each day. Also, we do not assume the RHS  $f$  corresponds to Gaussian white noise.

However, there are still some aspects that are neglected, which will be further studied to improve the current approach. For example, this approach is not tailored specifically to the evaluation criterion twCRPS. The performance might be further improved by putting the emphasis on the upper tail, according to the weight function in twCRPS. Also, reference weeks are just sampled randomly, which means some of the sampled reference fields might not be compatible with the known data. It will be helpful to study which Laplace field is more likely to present subject to the known data.

Thanks to the efficiency of our method, we were able to simulate a large number of temperature anomaly fields and compute local minimums directly. That is why extreme value theory methods were not used in our competition contribution. In the future, it will be studied to improve this model with EVT methods. For example, the sample size  $N$  can be largely reduced if EVT methods are used to model the

distributions of reference Laplace or gradient fields. This model can also be combined with the EVT method to further reduce its error on large temperature anomalies.

**Acknowledgements** We thank the organizing committee of the 11th international conference on Extreme Value Analysis, and Dr. Raphaël Huser for organizing the data challenge. Also, we thank the editors and anonymous reviewers whose suggestions helped improve and clarify this manuscript. Both authors are financially supported by China Scholarship Council for their PhD studies at TU Delft. The authors also acknowledge the support of their supervisors – they are D. Cheng’s supervisors Dr. Pasquale Cirillo, and Prof. dr. Frank Redig, and Z. Liu’s supervisors Dr. Eugeni L. Doubrovski, Prof. dr. Jo M.P. Geraedts, and Prof. Charlie C.L. Wang.

## Compliance with Ethical Standards

**Conflict of interests** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bacro, J.-N., Gaetan, C., Opitz, T., Toulemonde, G.: Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. *J. Am. Stat. Assoc.* **115**(530), 555–569 (2020). <https://doi.org/10.1080/01621459.2019.1617152>
- Chavez-Demoulin, V., Davison, A.C.: Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**(1), 207–222 (2005). <https://doi.org/10.1111/j.1467-9876.2005.00479.x>
- Davis, R.A., Klüppelberg, C., Steinkohl, C.: Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**(5), 791–819 (2013). <https://doi.org/10.1111/rssb.12012>
- Davison, A.C., Padoan, S.A., Ribatet, M.: Statistical modeling of spatial extremes. *Stat. Sci.* **27**(2), 161–186 (2012). <https://doi.org/10.1214/11-STS376>
- Godsil, C., Royle, G.F.: *Algebraic Graph Theory*. Springer-Verlag, New York (2001). <https://doi.org/10.1007/978-1-4613-0163-9>
- Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York (2009). <https://doi.org/10.1007/978-0-387-84858-7>
- Huser, R., Davison, A.C.: Space–time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(2), 439–461 (2014). <https://doi.org/10.1111/rssb.12035>
- Huser, R.: Editorial: EVA 2019 data competition on spatio-temporal prediction of Red Sea surface temperature extremes. *Extremes*. <https://doi.org/10.1007/s10687-019-00369-9> (2020)
- Knabner, P., Angerman, L.: *Numerical Methods for Elliptic and Parabolic Partial Differential Equations*. Springer-Verlag, New York (2003). <https://doi.org/10.1007/b97419>
- Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., Rue, H.: *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Chapman and Hall/CRC, Boca Raton (2018). <https://doi.org/10.1201/9780429031892>
- Lindgren, F., Rue, H., Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical*

- Society: Series B (Statistical Methodology) **73**(4), 423–498 (2011). <https://doi.org/10.1111/j.1467-9868.2011.00777.x>
- Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. *ACM Trans. Graph.* **22**(3), 313–318 (2003). <https://doi.org/10.1145/882262.882269>
- Pickands, J.: Statistical inference using extreme order statistics. *Ann. Stat.* **3**(1), 119–131 (1975). <https://doi.org/10.1214/aos/1176343003>
- Reich, B.J., Shaby, B.A.: A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat.* **6**(4), 1430–1451 (2012). <https://doi.org/10.1214/12-AOAS591>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.