Spatio-Temporal Weathering Predictions in the Sparse Data Regime with Gaussian Processes

Anonymous Author(s)
Affiliation
Address
email

Abstract

We investigate the problem of predicting the expected lifetime of a material in different climatic conditions from a few observations in sparsely located testing facilities. We propose a Spatio-Temporal adaptation of Gaussian Process Regression that takes full advantage of high-quality satellite data by performing an interpolation directly in the space of climatological time-series. We illustrate our approach by predicting gloss retention of industrial paint formulations. Furthermore, our model provides uncertainty that can guide decision-making and is applicable to a wide range of problems.

1 Introduction

Weathering refers to the in-service degradation of materials from the natural environment [1]. Being able to predict the service-life of materials under different climatic conditions is of critical importance to guarantee long-term safety, improve cost efficiency of design decisions during the construction phase and shorten development cycles for new products. However, it is usually a challenging problem, since several chemical degradation processes can occur simultaneously and are triggered by different environmental conditions. While there are research avenues focussing on chemical understanding [2, 3, 4] and empirical fitting [5, 6, 7], Machine Learning (ML) methods have gained popularity in the last decade, showing very promising effectiveness in modeling complex non-linear behaviours without the need to unravel the underlying physical processes [8, 9]. Within this context, in order to predict the material behaviour in untested locations, we seek a data-driven approach to model the relation between degradation over time and exposure location, i.e. a Spatio-Temporal (ST) problem. The problem is quite unique and differs from common ST forecasting [10] as no past observation is available at the new location. Furthermore, ST methods commonly rely on large datasets and perform interpolation between densely allocated spatial points. However, in weathering, this is hardly the case, since testing facilities tend to be few and isolated around the globe. In such cases of very sparse sampling, the spatial correlation is lost and the use of spatial interpolation or Deep Learning based ST methods is not possible [11, 12]. While geographical coordinates become uninformative, local climatic conditions still carry most of the information needed to identify material degradation at most locations [13]. This leads to our focus on climate representation.

Climate is itself a dynamical system and measurements mainly come as time-series (TS) observations. While previous works have relied on extracted features, such as mean values or fitted parameters [14, 15], we propose to retain the TS format to avoid any loss of information. The choice is also highly motivated by the recent developments in satellite data acquisitions which make high-quality climatology TS data available on a global scale. We then propose to represent each climate with a Multivariate Time Series (MTS), containing the time evolution of any relevant climatic variable. To predict degradation in new locations, we perform the interpolation in this highly informative climate space by adopting an ST adaptation of a Gaussian Process model (ST-GP).

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

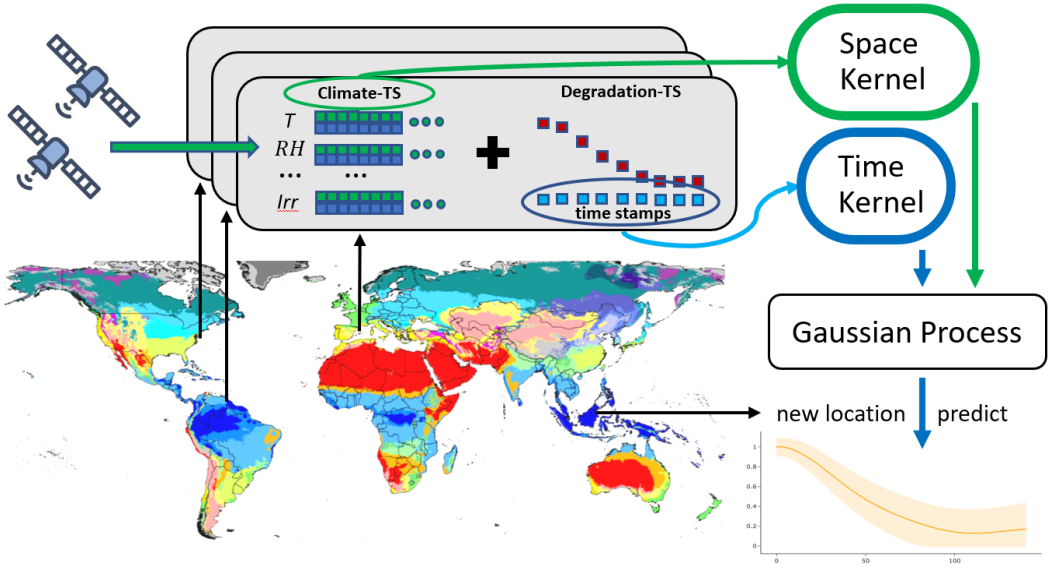

Figure 1: Illustration of the Spatio-Temporal Gaussian Process (ST-GP) framework. We pair two kernels, one space-based kernel using climatology data from satellites, and one time-based kernel using time-stamps of weathering data at testing facilities in archetypal climates. Their combination in a GP model enables prediction of performance in untested locations with uncertainty estimates.

The proposed framework benefits from great flexibility and leads us to the following main results:

- by pairing weathering data from testing facilities from archetypal climates with climatology satellite data, the model can predict performance in untested locations. This severely reduces the need for additional testing sites;
- by providing uncertainty estimates, the model also gives insight on where it is optimal to gather additional data;
- the model allows for extensions to incorporate other sources of variability such as chemical properties and can then support material discovery.

We begin with a presentation of the framework illustrated in Fig. 1, followed by quantitative and qualitative experimental evaluation.

1.1 Notation

Through the article, we indicate variables as lowercase (x); vectors and TS as bold lowercase ($\mathbf{x}$); matrices and MTS as bold uppercase $\mathbf{X}$. An index between square parenthesis $x[n]|_{n=1}^{N}$ indicate the n-th sample of a dataset and a subscript $x_t|_{t=1}^{T}$ the time stamp t.

2 Methodology

Consider a dataset of N_{loc} entities, where each (i) is a TS of length T_i which samples the degradation of the same quality variable y in a different location: $\mathbf{y}[i] \in \mathbb{R}^{T_i}|_{i=1}^{N_{loc}}$. In our framework, every degradation data entry $y_t[i]$ is associated with a time stamp (t), marking the exposure duration, and with an MTS of spatial descriptors ($\mathbf{C}[i]$), representing the climate at the exposure site:

$$\text{training points: } [\ y_t[i],\ t,\ \mathbf{C}[i]\ ] \quad \text{with:} \quad i = \{0, ..., N_{loc}\};\ t = \{0, ..., T_i\} \tag{1}$$

What remains is to model the dependence between the spatio-temporal inputs ($t, \mathbf{C}$) and the response y. We accomplish that by training a Gaussian Process (GP) [16], mainly motivated by the following considerations:

- GPs have remarkable modeling performances in the little data regime;
- GPs have been used in ST regression problems [17, 18] and geostatistics (*kriging*) [19], however, previous studies mainly rely on coordinates and spatial proximity;
- GPs are flexible with any kind of input, as long as a kernel function is provided for the data at hand. Furthermore, the closure of kernels under multiplication allows us to decouple the temporal and spatial dependence.

Following a common practice, we adopt a zero mean prior and do not introduce any prior knowledge into the model. Regarding the covariance function, we employ a decoupled spatio-temporal kernel composed of an RBF over time and the Global Alignment Kernel (GAK) [20] over the MTS spatial inputs:

$$k(t, t', \mathbf{C}, \mathbf{C}') \;=\; \theta_{\text{scale}} \underbrace{e^{\frac{|t-t'|^2}{2\sigma_{time}}}}_{k_{\text{time}}(t,\,t')} * \underbrace{\text{GAK}(\mathbf{C}, \mathbf{C}')}_{k_{\text{space}}(\mathbf{C},\,\mathbf{C}')} + \theta_{\text{noise}}\delta_{(t,t')} \tag{2}$$

where the hyperparameters are the variances θ_{scale} and θ_{noise}, the time-bandwidth σ_{time} and the GAK multiplicative factor parameter. The GAK was originally proposed for univariate series and can be extended to the multivariate case by averaging GAK on individual dimensions (independent GAK) or by considering alignments of vector-valued entries (dependent GAK) [21]. As our main claims will not be significantly affected by this choice, we present the results by adopting the independent GAK and show some comparisons with the alternative in Appendix B.

After having trained the ST-GP model, it is finally possible to predict the degradation $y(t, \bar{\mathbf{C}})$ for any t, by providing $\bar{\mathbf{C}}$ for a new location.

3 Application: weathering of paint formulations

In this section, we apply the developed framework to service-life predictions of paint formulations in untested locations. A toy-experiment with more abundant and publicly available data is presented in Appendix A.

3.1 Datasets

Climate data The data were obtained from the NASA Langley Research Center (LaRC) POWER Project's Climatology on 2022/01/18. The database contains data about surface solar energy fluxes and other meteorological quantities obtained through satellite systems and further reanalysis. For each point on a dense global-scale grid, climate data are provided in TS format with customizable time resolution. Further information, together with data and API, is available from the project website [22]. We choose the *climatology* TS format, consisting of 12 time-stamps with monthly values averaged over up to 30 years. We select the following 10 variables: mean temperature (C), relative humidity ($\%$), wind speed (m/s), dew/frost point (C), maximum temperature (C), cloud amount ($\%$), temperature range (C), precipitation ($mm/$day), all-sky surface shortwave and longwave irradiance (W/m^2). Each MTS then has the form $\mathbf{C} \in \mathbb{R}^{12\times 10}$.

Paints degradation data The degradation data have been privately provided by the Beckers group. They consist of TS data about the changes over time in a quality parameter, i.e. Gloss Retention (GR), for 37 different polyester-based paints systems (13 out-of-market qualities in 3 colour pigments: White, Red-Brown and Dark-Blue). Every GR value is obtained as the ratio between measured and initial gloss at $t = 0$ (details about gloss measures are available in [23]). For each one of the 37 formulations, the data come in two forms: TS obtained from long natural exposures (up to 139 months) in 10 different sites on a global scale and TS from 3 different controlled accelerated tests. The number of available time-stamps in each TS depends on the exposure location, ranging from 3 to 19, not necessarily equally spaced. A table with the availability in each location is shown in Appendix B.

3.2 Training on the same formulation

Consider one fixed formulation and the task of predicting its degradation curve in a new climate from exposure data of the same formulation in other locations. As the chemical variability is removed, we just need to take into account the climatic variability. For each fixed formulation, our data consist of $N_{loc} = 10$ degradation TS ($\mathbf{y}[i] \in \mathbb{R}^{T_i}$), each one paired with a climate MTS ($\mathbf{C}[i] \in \mathbb{R}^{12\times 10}$), collected from the POWER archive at each available testing site.

To internally test the interpolation performances, we withheld one degradation TS from one location and try to recover it by training the ST-GP model on the remaining $N_{loc} - 1$. To report the accuracy, we use the Mean Average Error (MAE) and the Mean Standardized Log Loss (MSLL), a model uncertainty aware score that computes a standardized negative probability of the target under the model; therefore is approximately zero for simple methods and negative for better methods. [24]. Both are computed at each withheld location, averaged across different fixed formulations and reported on the left-side in Tab. 1. Results are shown using the independent version of the GAK, with the bandwidth multiplicative factor set to $\sigma_{mul} = 2$. More details about the training and more insights on the variability with these choices are given in Appendix B. Some prediction examples are shown in Fig.2.

	same material GAK ind		expanded training GAK ind	
Withheld location	MAE	MSLL	MAE	MSLL
Allunga (Australia)	0.10	-0.83	0.06	-1.39
Miami (Florida)	0.11	-0.76	0.06	-1.31
Hainan (China)	0.13	-0.77	0.09	-0.90
Goa (India)	0.18	-0.78	0.08	-1.05
Kuala Lumpur (Malaysia)	0.20	2.17	0.08	-1.17
Singapore (Singapore)	0.22	1.79	0.09	-1.04
Vereeniging (South Africa)	0.10	-0.58	0.07	-1.23
Montbrison (France)	0.07	-1.24	0.05	-1.40
Liverpool (England)	0.08	-1.00	0.08	-0.89
Bohus Malmon (Sweden)	0.11	-0.89	0.08	-1.07

Table 1: The first two numerical columns contain the MAE and MSLL between the data in the withheld location and the model predictions; the last two columns show the same scores when the training data is expanded with similar formulations in the same climate.

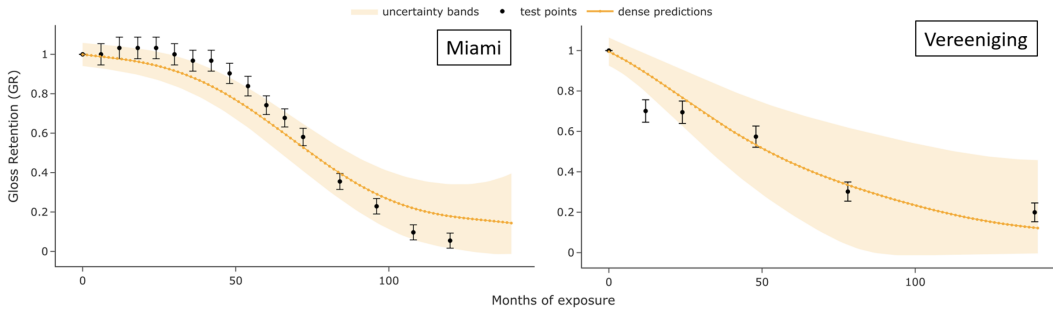

Figure 2: Degradation predictions in two withheld locations. The ST-GP framework presented here allows prediction of degradation that is quantitatively consistent with the withheld testing facility data.

Given the complexity of the problem, it is a success to find most average errors around $\Delta GR \sim 0.1$ and negative MSLL. Of particular interest is the interpolation accuracy in the arid climate of South Africa, climatologically far from all the other locations. On the contrary, we would expect small errors at the cities of Singapore and Kuala Lumpur, as one should be very informative in predicting the other. We observe instead a high and positive MSLL. In this case, the gloss data in the city of

Kuala Lumpur are mainly influenced by the heavy accumulation of dust and dirt (as opposed to Singapore), for which appropriate descriptors have not been included. These results indicate that climatic agents do not always constitute the only source of degradation and different effects might be important too. These can easily be incorporated into the framework by expanding **C** and reserved for a future study.

As a parallel experiment, we redirect predictions to new locations where we do not have data. Instead of the whole service-life of the material, we focus now on one particular time stamp, i.e. 10 years of exposure, but extend the predictions to a dense grid in a geographical region (which is possible thanks to the high resolution of the satellite climate data). Even if we do not have testing data, this represents a qualitative validation as we can check if the predictions reflect our intuition. In general terms, one can expect harsher degradation in hot and aggressive climates, while steadier performances in more rigid ones. As it can be seen in Fig. 3, the degradation map in Italy indeed reflects many features of

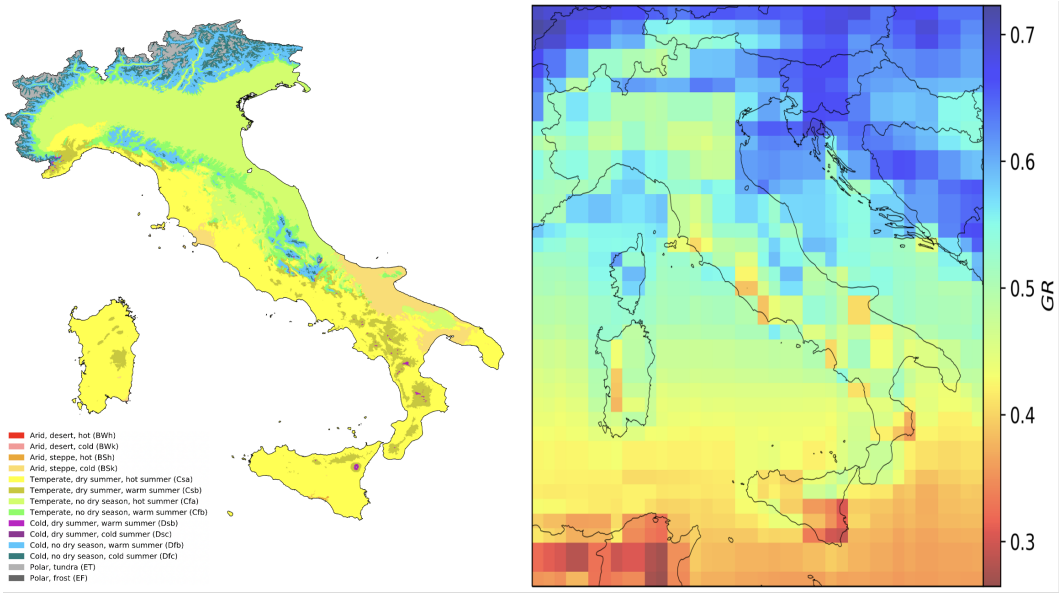

Figure 3: Left) Köppen Geiger map of Italy. Right) Predictions of GR degradation after 10 years of exposure across Italy, for formulation #37.

the corresponding Köppen Geiger map [25]. This is a stunning result, considering that the training set is composed of only 10 sparse climates. On top of that, only a few locations contain data for up to 10 years. Of particular interest is the relatively harsh degradation predicted on the Alps, which, despite the rigid climate, can be expected as a consequence of the high altitude and the high level of UV radiation. Additional studies will be conducted to validate such considerations.

As we show in Fig. 4, it is also possible to observe the geographical dependence of the model reliability by plotting the uncertainties of the ST-GP on a global scale. This comes with many advantages, such as guiding decision making, optimal gathering of additional data and support warranties.

3.3 Training on more formulations

During the previous sections, we have restricted our focus to climatic variability only. In a more general case, the flexibility of composing different kernels allows for extending the variability over different entities. In our case, it is possible to extend the variability over different chemicals by considering the 3 accelerated test data. We use them as MTS descriptors that capture dynamic properties of each formulation rather than the location they are exposed at. In practice, each formulation exposed in a specific climate now corresponds to an unique pairing of climate MTS ($\mathbf{C}_{clm}$) and accelerated MTS ($\mathbf{C}_{acc}$). The independent GAK averaged across all variables contained in $\mathbf{C}_{clm}$ and $\mathbf{C}_{acc}$ is finally able to distinguish between all 37×10 instances.

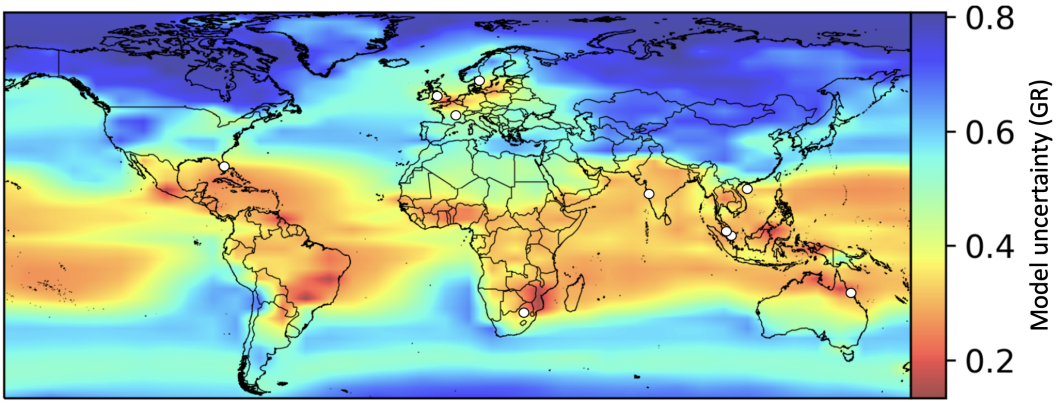

Figure 4: Uncertainties (2σ) on model predictions in correspondence of 10 years of exposure. White dots represent locations containing degradation data.

To demonstrate that chemical similarity can be captured in such a way, we consider a similar cross-validation experiment to Sec. 3.2 but expand the training data to all other formulations exposed in all climates, including the target one. Since similar materials degrade similarly in the same climate, the model may now have access to formulations with similar behaviour to what we try to predict. What we are testing here is the ability of the model to give more importance to those data. If the similarity is captured correctly, we expect this task to be easier, with higher overall accuracy. The experimental results are reported in the last column of Tab. 1 and confirm this hypothesis. Note that the missing descriptors for Kuala Lumpur and Singapore are much less relevant in this scenario. We also report that the accuracies can be further refined by considering an additional kernel term that takes into account the variability over static chemical properties, e.g. the T_g of the formulation. However, discussing specific properties goes beyond the scope of this note and the relative results will then not be deepened.

4 Conclusions and future directions

In this work, we have introduced a Spatio-Temporal adaptation of Gaussian Process Regression to predict the in-service degradation of materials under different weathering conditions. Our approach finds its most optimal applicability in the gap between advanced ST-ML methods, of which the usage is discouraged because of data sparsity, and knowledge-based methods that cannot capture the complex dependence between climate and chemical degradation. Both quantitative and qualitative validations have been presented.

Regarding future works, we consider of critical importance the inclusion of additional validation arguments, either by collecting more data or by studying the relevance of each climate variable in the predictions. As an additional important follow-up to this work, we will also consider including more domain knowledge into the framework, e.g. additional spatial descriptors (air quality, chemical properties, etc.) or monotonicity constraints in the GP. As opposed to climatological monthly values, daily data collected during the exposure (also easily available through [22]) might constitute even more informative climatic descriptors. To deal with the poor scalability of the GAK, a novel kernel should be paired with this choice to process very long TS.

Acknowledgments and Disclosure of Funding

We would like to acknowledge useful discussions with James Maxted from the Beckers Group. This research is supported by the Beckers Group: `https://www.beckers-group.com`. Climate data were obtained from the NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science/Applied Science Program.

References

[1] George Wypych. *Handbook of material weathering*. Elsevier, 2018.

[2] David R Bauer, Jonathan W Martin, et al. *Service life prediction of organic coatings*. ACS Publications, 1999.

[3] Andrew Hulme and Jenny Cooper. Life prediction of polymers for industry. *Sealing Technology*, 2012(9):8–12, 2012.

[4] FP La Mantia, M Morreale, L Botta, MC Mistretta, M Ceraulo, and R Scaffaro. Degradation of polymer blends: A brief review. *Polymer Degradation and Stability*, 145:79–92, 2017.

[5] Susanne Kahlen, GM Wallner, and RW Lang. Aging behavior and lifetime modeling for polycarbonate. *Solar Energy*, 84(5):755–762, 2010.

[6] Emad Yousif and Raghad Haddad. Photodegradation and photostabilization of polymers, especially polystyrene. *SpringerPlus*, 2(1):1–32, 2013.

[7] Bronwyn Laycock, Melissa Nikolić, John M Colwell, Emilie Gauthier, Peter Halley, Steven Bottle, and Graeme George. Lifetime prediction of biodegradable polymers. *Progress in Polymer Science*, 71:144–189, 2017.

[8] Will Nash, Tom Drummond, and Nick Birbilis. A review of deep learning in the study of materials degradation. *npj Materials Degradation*, 2(1):1–12, 2018.

[9] Jiangfeng An, Duncheng Peng, Xuejie Zhou, Jun Wu, and Penghua Zheng. Service-life study of polycarbonate outdoors using python with incomplete data. *Modelling and Simulation in Engineering*, 2020, 2020.

[10] Xingjian Shi and Dit-Yan Yeung. Machine learning for spatiotemporal sequence forecasting: A survey. *arXiv preprint arXiv:1808.06865*, 2018.

[11] Senzhang Wang, Jiannong Cao, and Philip Yu. Deep learning for spatio-temporal data mining: A survey. *IEEE transactions on knowledge and data engineering*, 2020.

[12] Federico Amato, Fabian Guignard, Sylvain Robert, and Mikhail Kanevski. A novel framework for spatio-temporal prediction of environmental data using deep learning. *Scientific reports*, 10(1):1–11, 2020.

[13] Jin Gao, Chao Li, Zhen Lv, Rui Wang, Dequan Wu, and Xiaogang Li. Correlation between the surface aging of acrylic polyurethane coatings and environmental factors. *Progress in Organic Coatings*, 132:362–369, 2019.

[14] Dequan Wu, Dawei Zhang, Shaopeng Liu, Zhihui Jin, Thee Chowwanonthapunya, Jin Gao, and Xiaogang Li. Prediction of polycarbonate degradation in natural atmospheric environment of china based on bp-ann model with screened environmental factors. *Chemical Engineering Journal*, 399:125878, 2020.

[15] Han Liu, Mingyong Zhou, Yuli Zhou, Shan Wang, Guangxian Li, Long Jiang, and Yi Dan. Aging life prediction system of polymer outdoors constructed by ann. 1. lifetime prediction for polycarbonate. *Polymer degradation and stability*, 105:218–236, 2014.

[16] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

[17] Seth R Flaxman. *Machine learning in space and time*. PhD thesis, Ph. D. thesis, Carnegie Mellon University, 2015.

[18] Ransalu Senanayake, Simon O'Callaghan, and Fabio Ramos. Predicting spatio-temporal propagation of seasonal influenza using variational gaussian process regression. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

[19] Margaret A Oliver and Richard Webster. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System*, 4(3):313–332, 1990.

[20] Marco Cuturi. Fast global alignment kernels. In *ICML*, volume 2011, pages 929–936, 2011.

[21] Ahmed Shifaz, Charlotte Pelletier, Francois Petitjean, and Geoffrey I Webb. Elastic similarity measures for multivariate time series classification. *arXiv preprint arXiv:2102.10231*, 2021.

[22] POWER Project. `https://power.larc.nasa.gov/`.

[23] ISO 2813:2004. Paints and varnishes — Determination of gloss value at 20°, 60° and 85°. Standard, International Organization for Standardization, 2004.

[24] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[25] Hylke E Beck, Niklaus E Zimmermann, Tim R McVicar, Noemi Vergopolan, Alexis Berg, and Eric F Wood. Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific data*, 5(1):1–12, 2018.

[26] Romain Tavenard, Johann Faouzi, Gilles Vandewiele, Felix Divo, Guillaume Androz, Chester Holtz, Marie Payne, Roman Yurchak, Marc Rußwurm, Kushal Kolar, and Eli Woods. Tslearn, a machine learning toolkit for time series data. *Journal of Machine Learning Research*, 21(118):1–6, 2020.

[27] GPy. GPy: A gaussian process framework in python. `http://github.com/SheffieldML/GPy`, since 2012.

[28] Jacob R Gardner, Geoff Pleiss, David Bindel, Kilian Q Weinberger, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Advances in Neural Information Processing Systems*, 2018.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] framework for spatio-temporal predictions with sparse data, application to weathering, uncertainty maps to guide decision making.
   (b) Did you describe the limitations of your work? [Yes] Failures have been observed when the dominant factors are not included in the spatial MTS descriptors. This require domain knowledge in choosing the right variables.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] The degradation data and code are proprietary. Appendix A provides a toy-experiment employing the framework with a publicly available dataset. Link to the climate database has been given in Sec. 3.1.

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] Details for reproducibility have been included in the supplementary material for both experiments.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] In appendix, we report different results for different choices of the only unoptimized hyperparameter.

(d) Did you include the total amount of computing and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Execution time can be found in the supplementary material.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? [Yes] All assets are referenced in Appendix A.

(b) Did you mention the license of the assets? [N/A]

(c) Did you include any new assets either in the supplemental material or as a URL? [N/A] We do not propose new assets

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] Consent of data usage have been given by the Beckers group.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Toy-experiment: climate of world capitals

We present here a test of the ST-GP in a controlled experiment with climate data only. This serves as a way of supporting the weathering predictions in Sec. 3.2, furthermore, it provides results in more data abundant context and all within a fully publicly available dataset.

For each of the 235 capitals in the world, we interrogate the NASA climate dataset and collect the temporal evolution of the same 10 variables as in Sec. 3.1. The dataset we have built then consists of 235 MTS $\in \mathbb{R}^{12\times 10}$ distributed across the globe. We also standardize to zero mean and unit standard deviation within each dimension. To simulate our regression framework, we choose one variable as response y and the remaining as spatial descriptors $\mathbf{C}$. Because of its correlation with other variables, its relatively smooth evolution and variety of shapes, we choose 'mean temperature' as our response variable. If we now remove the response from one instance, the problem of retrieving its evolution from observations in other locations is equivalent to the regression problem of Sec. 3.2. Predictions of this kind can be iterated by hiding the response from one location at a time.

For the GAK bandwidth, we set its value to eight times the median distance of the MTS in the dataset, scaled by the square root of the median length. We have observed that equally high multiples work similarly well (e.g. 5 or 10) and better than lower multiples. Regarding all the other parameters, we set them to the average result of a preliminary Leave One Out Cross Validation (LOOCV) cycle by training on the 20 nearest locations in the kernel space and minimizing the negative Type-II MLE (marginal log-likelihood). A second cycle, with all $N_{loc} - 1$ instances included in the training, concludes with an average $R^2 = 0.962$ and average MAE $= 0.020$. The experiment runs in $\sim 2.5\,h$ with an AMD Ryzen™ 5 3600X processor. Some examples of predictions are shown in Fig. 5. With very few exceptions, the model was always able to predict the correct shape. Less accurate

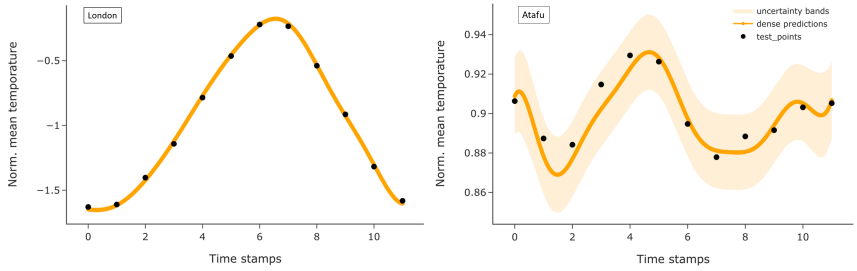

Figure 5: Examples of mean temperature predictions (orange curve and uncertainty bands) in correspondence of withheld testing data for 2 different capitals.

predictions are always paired with higher uncertainties.

A similar but more ambitious experiment can be set in the case of predicting daily climatic data. For the first 50 days of 2018, we performed a LOOCV cycle and report a lower average accuracy of MAE $= 0.066$. This is expected, as the higher noise level decorrelates nearby locations in the climate space. One successful example, for the first 200 days in the city of London, is shown in Fig. 6.

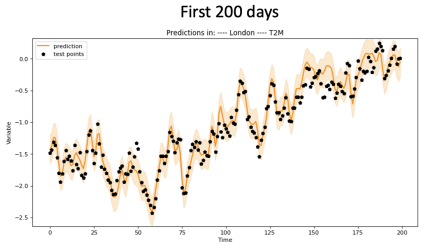

Figure 6: Mean daily temperature predictions in the city of London by interpolating nearby climates.

B Paint formulations weathering

B.1 Data table

Fig. 7 shows the availability of data within each location. The letter 'G' indicates that a GR measure is available for the corresponding location and time-stamp (top row).

	G = Gloss measurements / Gloss retention																						
Exposure Sites / Months	Original	6	9	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120	139
Allunga	G	G		G	G	G		G		G		G		G		G		G		G		G	
Miami	G	G		G	G	G	G	G	G	G	G	G	G	G		G		G		G		G	
Hainan	G	G		G	G	G	G																
Goa	G					G										G							
Kuala Lumpur	G	G	G	G	G	G	G	G	G	G				G		G		G	G	G	G	G	
Singapore	G	G		G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Vereeniging	G			G		G				G					G								G
Montbrison	G	G		G		G																	
Liverpool	G			G				G														G	
Bohus Malmon	G			G		G		G						G				G					

Figure 7: Data availability in each location.

B.2 Training details

For what regards the GAK bandwidth, we set its value to σ_{mul} times the median distance of the MTS in the dataset, scaled by the square root of the median length. Regarding all the other parameters, we set them by minimizing the negative Type-II MLE (marginal log-likelihood) in each iteration of the LOOCV cycle. The only bounded parameter in the minimization is the bandwidth of the time kernel: $\sigma_{time} \in [20, 150]$. The multiplicative factor σ_{mul} is then our only fixed hyperparameter. Tab. 2 contains several experiments with the two different GAK variants and three values of σ_{mul}. For the dependent GAK, the variability is negligible, except in Bohus Malmon. For the independent GAK, better accuracies tend to be found in the correspondence of lower multiples. In general, the claims in the article are not significantly affected by the choice of this parameter.

	GA dep			GA ind		
Withheld location	$\sigma_{mul} = 2$	$\sigma_{mul} = 5$	$\sigma_{mul} = 8$	$\sigma_{mul} = 2$	$\sigma_{mul} = 5$	$\sigma_{mul} = 8$
Allunga	0.077	0.072	0.078	0.101	0.125	0.136
Miami	0.110	0.104	0.101	0.114	0.139	0.153
Hainan	0.128	0.117	0.109	0.133	0.139	0.133
Goa	0.116	0.151	0.151	0.178	0.213	0.210
Kuala Lumpur	0.196	0.186	0.184	0.201	0.216	0.219
Singapore	0.201	0.197	0.199	0.216	0.228	0.231
Vereeniging	0.133	0.133	0.136	0.101	0.119	0.141
Montbrison	0.047	0.050	0.056	0.072	0.060	0.056
Liverpool	0.108	0.114	0.102	0.082	0.089	0.092
Bohus Malmon	0.135	0.128	0.099	0.108	0.113	0.113

Table 2: MAE in correspondence of each withheld location, averaged across all 37 formulations. Results are given for different values of the multiplicative factor in the GAK.

Each experiment for a given configuration runs in less than 10 minutes with an AMD Ryzen™ 5 3600X processor.

B.3 Implementation details

This experiment, as well as the one in Appendix A, has been implemented in Python 3.9; for the GAK we use the implementation provided by the package *tslearn* [26] while the GP training has been done both within *GPy* [27] and *GPyTorch* [28], producing compatible results.