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

Giovanni De Felice
Dep. of Computer Science
University of Liverpool
g.de-felice@liverpool.ac.uk

Vladimir V. Gusev
Dep. of Computer Science
University of Liverpool

John Y. Goulermas*
Dep. of Computer Science
University of Liverpool

Michael W. Gaultois
Dep. of Chemistry
University of Liverpool

Matthew J. Rosseinsky
Dep. of Chemistry
University of Liverpool

Catherine V. Gauvin
Beckers Group

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, known as ST kriging [10], is quite unique and differs from common ST forecasting [11] as no past observation is available at the new location. Furthermore, the most popular 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 [12, 13]. While geographical coordinates become uninformative, local climatic conditions still carry most of the information needed to identify material degradation at most locations [14] and can be used as contextual features [15]. This leads to our focus on climate representation.

*Deceased May 2022

NeurIPS 2022 AI for Science Workshop.

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 [16, 17], 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).

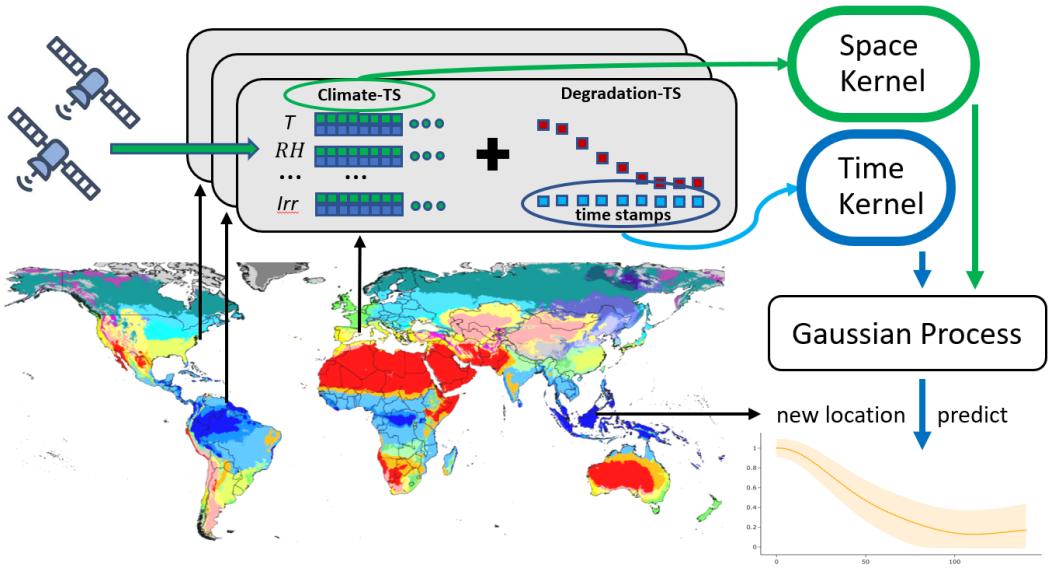

Figure 1: Illustration of the Spatio-Temporal Gaussian Process (ST-GP) framework. We pair two kernels, one space-based kernel using climatology data from satellites, and one time-based kernel using time-stamps of weathering data at testing facilities in archetypal climates. Their combination in a GP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

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

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) [18], mainly motivated by the following considerations:

- GPs have remarkable modeling performances in the little data regime;
- GPs have been used in ST regression problems [19, 20, 21, 22] and geostatistics (*kriging*) [23], 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 [24].

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) [25] over the MTS spatial inputs:

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

where the hyperparameters are the variances θ_{scale} and θ_{noise}, the time-bandwidth σ_{time} and the GAK bandwidth multiplicative factor σ_{mul}. Other kernels have been proposed for MTS objects [26, 27, 28]. However, they either require strong assumptions about the underlying distribution of the data or an ensemble strategy, which do not fit our little data regime. On top of that, the GAK can potentially handle series of different lengths, which can come in handy if contextual features are sampled irregularly over locations. As downsides, it is computationally very expensive and does not model the interdependencies between variables. The GAK was, if fact, 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) [29]. 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. For the design choice in Eq. 2, we preferred the multiplication (AND) between temporal and spatial inputs over the summation (OR) as we do not want to relate similar time-stamps if they belong to very different climates. In fact, we have observed that using summation leads to worse interpolation accuracies.

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 [30]. 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 a global supplier of coil coatings and industrial paints. 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 [31]). 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. [32]. 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 $\sigma_{mul} = 2$. Some prediction examples are shown in Fig.2. More details about the training and more insights on the variability with these choices are given in Appendix B, which also contains a brief 'ablation' study to deepen the relation between performances and the choice of the climate variables.

	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.

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

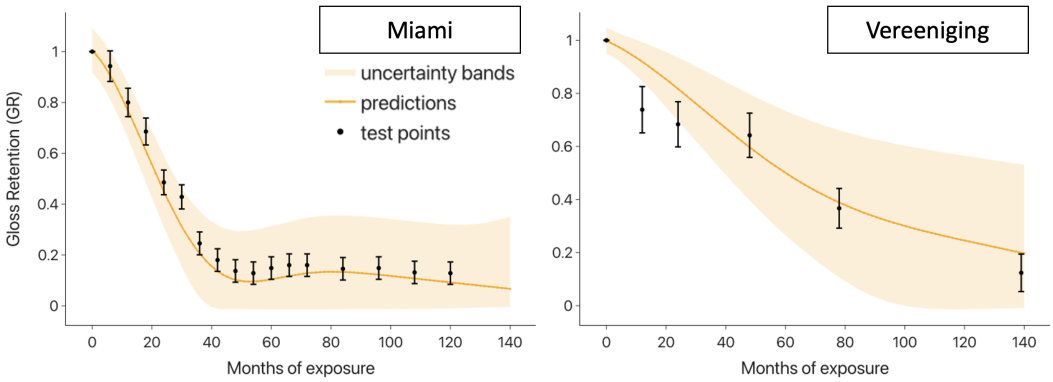

Figure 2: Degradation predictions in two withheld locations. The presented ST-GP framework allows for prediction of degradation that is quantitatively consistent with withheld testing facility data.

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 confirm 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

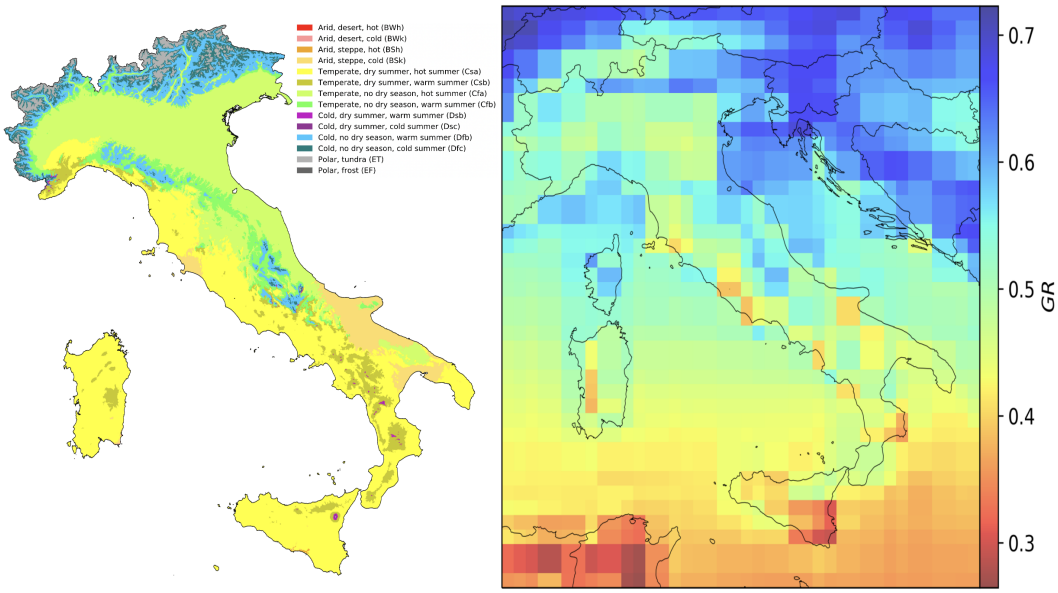

Figure 3: Degradation predictions in an extended geographical space. For formulation #37, the presented ST-GP framework produces a degradation map after 10 years of exposure (right) that qualitatively reflects the corresponding Köppen Geiger map (left).

the corresponding Köppen Geiger map [33]. 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. Note how the uncertainties are contained not only in presence of the available testing sites. This comes with many advantages, such as guiding decision making, optimal gathering of additional data and support warranties.

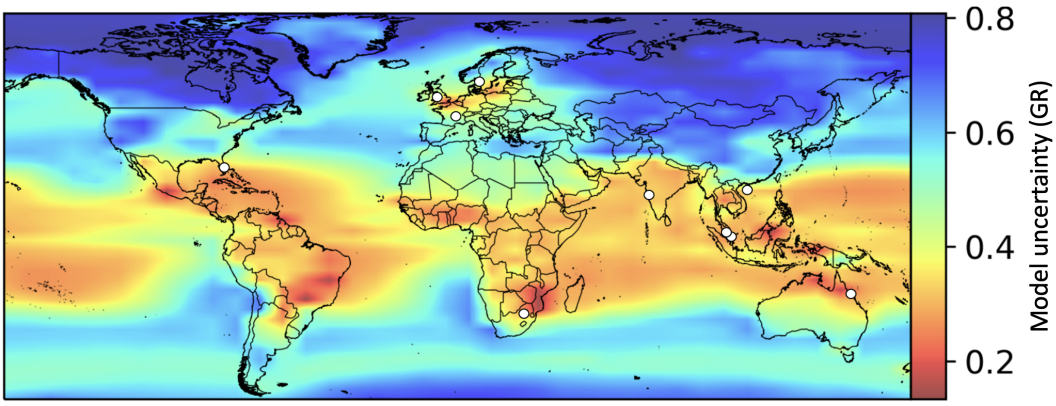

Figure 4: Model uncertainty map. The presented ST-GP framework allows for building decision-guiding uncertainty maps, of which an example is here given for formulation #37, at 2σ, after 10 years of exposure. Note how the uncertainties are contained not only at the available testing sites (white dots).

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.

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 glass-transition temperature of the formulation. However, discussing specific properties goes beyond the scope of this note and the relative results will 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. We presented both quantitative and qualitative validations. The model does not make any assumption on the underlying physical processes, making it suitable for other applications, such as plastic degradation or green-energy output.

Regarding future works, we consider of critical importance the inclusion of additional validation arguments, most importantly accompanied by collecting more data. 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 [30]) might constitute even more informative climatic descriptors as well as satellite-image data. However, this should also be accompanied by an enrichment of data and a kernel to deal with either long TS or images.

Acknowledgments and Disclosure of Funding

We would like to acknowledge useful discussions with James Maxted and other members of 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] Yuankai Wu, Dingyi Zhuang, Aurelie Labbe, and Lijun Sun. Inductive graph neural networks for spatiotemporal kriging. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5):4478–4485, May 2021.

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

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

[13] 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.

[14] 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.

[15] Zhe Jiang. A survey on spatial prediction methods. *IEEE Transactions on Knowledge and Data Engineering*, 31(9):1645–1664, 2018.

[16] 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.

[17] 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.

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

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

[20] 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.

[21] Mu Niu, Zhenwen Dai, Neil Lawrence, and Kolja Becker. Spatio-temporal gaussian processes modeling of dynamical systems in systems biology. *arXiv preprint arXiv:1610.05163*, 2016.

[22] Danil Kuzin, Olga Isupova, and Lyudmila Mihaylova. Spatio-temporal structured sparse regression with hierarchical gaussian process priors. *IEEE Transactions on Signal Processing*, 66(17):4598–4611, 2018.

[23] 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.

[24] Shiwei Lan. Learning temporal evolution of spatial dependence with generalized spatiotemporal gaussian process models. *Journal of Machine Learning Research*, 23(259):1–53, 2022.

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

[26] Tony Jebara, Risi Kondor, and Andrew Howard. Probability product kernels. *The Journal of Machine Learning Research*, 5:819–844, 2004.

[27] Mustafa Gokce Baydogan and George Runger. Time series representation and similarity based on local autopatterns. *Data Mining and Knowledge Discovery*, 30(2):476–509, 2016.

[28] Karl Øyvind Mikalsen, Filippo Maria Bianchi, Cristina Soguero-Ruiz, and Robert Jenssen. Time series cluster kernel for learning similarities between multivariate time series with missing data. *Pattern Recognition*, 76:569–581, 2018.

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

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

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

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

[33] 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.

[34] 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.

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

[36] 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.

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 capital cities 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. With very few exceptions, the model was always able

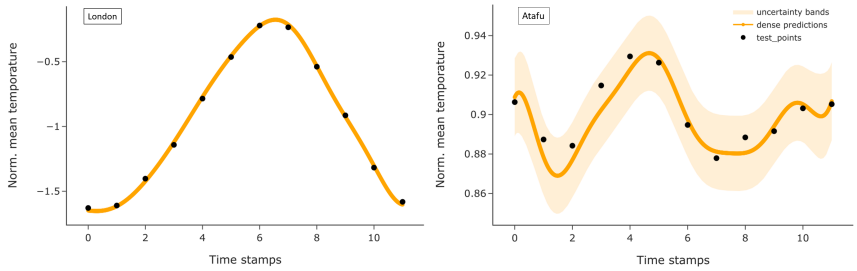

Figure 5: Mean temperature predictions in correspondence of withheld testing data for 2 different capitals.

to predict the correct shape. Less accurate predictions are always paired with higher uncertainties. Some examples of predictions are shown in Fig. 5.

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.

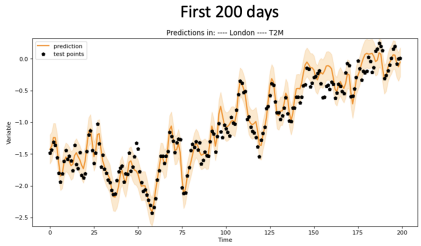

Figure 6: Mean daily temperature predictions in the city of London by interpolating nearby climates.

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 of the paint degradation dataset.

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* [34] while the GP training has been done both within *GPy* [35] and *GPyTorch* [36], producing compatible results.

B.4 Ablation study

The 10 climate variables of Sec. 3.1 were selected on the basis of expertise on GR degradation. In this section, we deepen the influence of individual climate variables on the predictions accuracy. We conduct experiment analogues to Sec. 3.2 but, instead of averaging the individual kernels over different climate variables (GAK ind), we limit the calculation to a single variable (Tab. 3) or a subset of them (Tab. 4). For simplicity, we do not show results for withheld locations that contains only short-term exposure data (Hainan, Montbrison) or contains ≤ 3 points (Goa). Also we omit results for Kuala Lumpur and Singapore as the errors are dominated by the lack of appropriate descriptors (as discussed in Sec. 3.2).

Individual variables Tab. 3 contains the MAE in correspondence of withheld locations when only a single climate variable is used to calculate the space-based kernel (k_{space}). The accuracy by using all 10 variables is also reported for convenience. We notice that, in general, the combination of all variables tend to outperform individuals. One exception is given by only using 'precipitation' data, which leads to interestingly good accuracies. An interpretation might be that 'precipitation' data carry much information as they have a direct impact on other variables such as temperature, humidity and cloud cover. This might be enough to distinguish between such different climates at this level of precision. However, it is hard to believe that this variable alone would describe all climate variability at a regional level. As expected, Fig. 8 (left) shows how this setting does not produce a qualitatively appealing degradation map as we observed in Fig. 3.

	GA ind							
Withheld location	T	RH	WS	T_{dew}	CA	P	[L,S irr]	all
Allunga	0.19	0.19	0.13	0.10	0.09	0.10	0.15	0.10
Miami	0.15	0.17	0.25	0.08	0.15	0.10	0.11	0.11
Vereeniging	0.23	0.22	0.17	0.22	0.14	0.10	0.19	0.10
Liverpool	0.13	0.09	0.20	0.09	0.17	0.07	0.11	0.08
Bohus Malmon	0.16	0.13	0.26	0.16	0.26	0.09	0.14	0.11

Table 3: The table contains the MAE between the data in the withheld location and the model predictions, when only one of the following individual climatic variable is used as spatial descriptor: mean temperature (T), relative humidity (RH), wind-speed (WS), dew-point (T_{dew}), cloud amount (CA), precipitations (P), all-sky surface shortwave and longwave irradiance (L,S irr).

	GA ind		
Withheld location	$[T, RH, WS]$	all - $[T_{range}, T_{max}]$	all
Allunga	0.12	0.08	0.10
Miami	0.16	0.11	0.11
Vereeniging	0.20	0.10	0.10
Liverpool	0.12	0.08	0.08
Bohus Malmon	0.15	0.11	0.11

Table 4: The table contains the MAE between the data in the withheld location and the model predictions, when a subset of climatic variables is used as spatial descriptor. The first column indicate the subset of mean temperature (T), relative humidity (RH) and wind-speed (WS); the second column indicate the set of all variables except for temperature range (T_{range}) and maximum temperature (T_{max}).

Subsets of variables The first numerical column of Tab. 4 shows an example of how extending to a subset of three variables leads to an overall better accuracy with respect to the individuals (which lack generality). By adding more, we have observed that the accuracies converge towards better results until they match the setting with the complete set of variables.

Correlated variables The complete set contains three very correlated variables, i.e. 'mean temperature', 'maximum temperature' and 'temperature range'. This was done with the intention of modeling

heavy stresses on the material caused by temperature jumps. However, as it can be observed in the second numerical column of Tab. 4, removing two of them does not impact the accuracy. On top of that, the degradation map (Fig. 8, right) still looks reasonable (yet not identical). This suggests that the model might not be sensitive to this effect, most probably due to the scarcity of data.

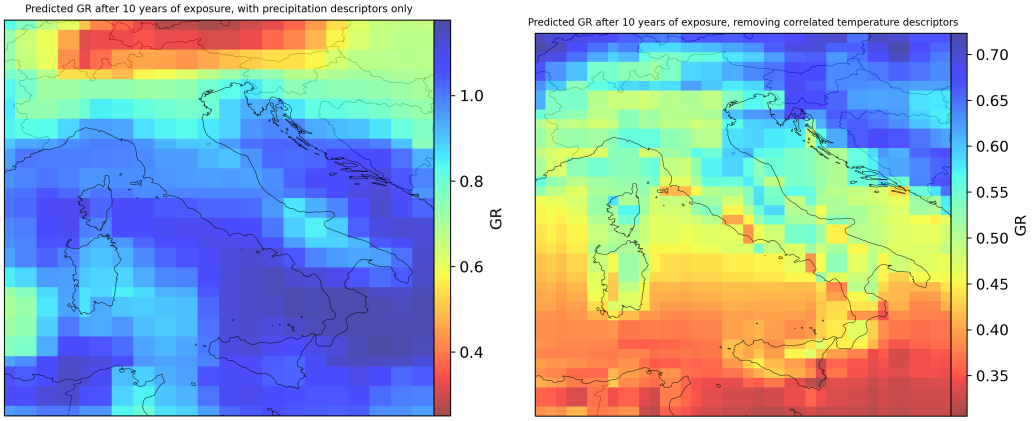

Figure 8: Degradation predictions in an extended geographical space with two different settings. When only precipitation data are used as spatial descriptors, the model produces an unsatisfactory degradation map (left). When temperature range (T_{range}) and maximum temperature (T_{max}) are stripped from the spatial descriptors, the model still outputs a satisfactory degradation map (right) that reflect geographical features and intuition.