



Modeling the R -ratio and hadronic contributions to $g - 2$ with a Treed Gaussian process

Andrew Fowlie^{1,a}, Qiao Li^{2,b}

¹ Department of Physics, School of Mathematics and Physics, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China

² Department of Physics, Institute of Theoretical Physics, Nanjing Normal University, Nanjing 210023, Jiangsu, China

Received: 4 July 2023 / Accepted: 4 October 2023 / Published online: 19 October 2023
© The Author(s) 2023

Abstract The BNL and FNAL measurements of the anomalous magnetic moment of the muon disagree with the Standard Model (SM) prediction by more than 4σ . The hadronic vacuum polarization (HVP) contributions are the dominant source of uncertainty in the SM prediction. There are, however, tensions between different estimates of the HVP contributions, including data-driven estimates based on measurements of the R -ratio. To investigate that tension, we modeled the unknown R -ratio as a function of CM energy with a treed Gaussian process (TGP). This is a principled and general method grounded in data-science that allows complete uncertainty quantification and automatically balances over- and under-fitting to noisy data. Our tool yields exploratory results are similar to previous ones and we find no indication that the R -ratio was previously mismodeled. Whilst we advance some aspects of modeling the R -ratio and develop new tools for doing so, a competitive estimate of the HVP contributions requires domain-specific expertise and a carefully curated database of measurements ([github](https://github.com/qiao688/TGP_for_g-2), https://github.com/qiao688/TGP_for_g-2).

1 Introduction

The final measurement of the anomalous magnetic moment of the muon at the Brookhaven National Laboratory (BNL) E821 experiment [1] differed from the Standard Model (SM) prediction by 3.7σ . This discrepancy was replicated by the Fermi National Accelerator Laboratory (FNAL) E989 measurement [2]. The combined measurement from BNL and FNAL,

$$a_\mu = 116\,592\,061(41) \times 10^{-11}, \quad (1)$$

^a e-mail: andrew.fowlie@xjtlu.edu.cn (corresponding author)

^b e-mail: qiaoli@njnu.edu.cn

deviates from the SM theory prediction by 4.2σ , motivating the possibility of physics beyond the SM [3] as well as scrutiny of the SM prediction [4].

The SM prediction for a_μ incorporates contributions from quantum electrodynamics (QED), electroweak interactions (EW), and hadronic effects [5]. While QED and EW contributions can be calculated with high precision using perturbation theory and are well-controlled [6], hadronic contributions are harder to compute and are the largest source of uncertainty in predictions for a_μ . The hadronic contribution can itself be decomposed into the following parts:

$$a_\mu^{\text{had}} = a_\mu^{\text{HVP}} + a_\mu^{\text{LbL}}, \quad (2)$$

where a_μ^{HVP} is the hadronic vacuum polarization (HVP) contribution and a_μ^{LbL} is the light-by-light scattering contribution. In this work we focus on the leading-order (LO) contribution to HVP. This is particularly challenging and the dominant source of uncertainty, making up about 80% of the total. The two most popular computational methods are lattice QCD and estimates from dispersion integrals and cross-section data.

Lattice methods calculate the HVP contribution by discretizing spacetime and performing a weighted integral of relevant functions over Euclidean time. This requires significant computational resources and currently cannot match the nominal precision achieved by data-driven methods. Data-driven methods use data from, for example, the KLOE [7], BaBar [8–11], SND [12] and CMD-3 [13, 14] experiments to estimate the R -ratio,

$$R(s) \equiv \frac{\sigma^0(e^+e^- \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)}. \quad (3)$$

This is a function of the center-of-mass (CM) energy, \sqrt{s} .¹ From the estimate of the R -ratio, the leading-order (LO) HVP contributions can be computed through the dispersion integral,

$$a_{\mu}^{\text{HVP}} = \frac{\alpha^2}{3\pi^2} \int_{m_{\pi}^2}^{\infty} \frac{K(s)R(s)}{s} ds \quad (4)$$

where $K(s)$ is the QED kernel [15, 16]. Hadronic contributions to the effective electromagnetic coupling constant at the Z boson mass can be computed in a similar way through the dispersion relationship

$$\Delta\alpha_{\text{had}} = \frac{\alpha M_Z^2}{3\pi} \int_{m_{\pi}^2}^{\infty} \frac{R(s)}{s(M_Z^2 - s)} ds \quad (5)$$

where f represents the principal-value prescription. There is growing tension between results from the two approaches [17, 18]. A recent lattice QCD calculation found [19]

$$a_{\mu}^{\text{HVP}} = 707.5 (5.5) \times 10^{-10} \quad (6)$$

whereas a conservative combination of data-driven estimates yielded [4]

$$a_{\mu}^{\text{HVP}} = 693.1 (4.0) \times 10^{-10}. \quad (7)$$

The lattice result was at least partly corroborated by other recent lattice computations [20–22] and, moreover, data-driven estimates using hadronic τ -decays are close to lattice results [23]. The most recent measurement of the 2π final state at CDM-3 [24] compounded the mystery, as it conflicts with older measurements including CDM-2 [25].

With these issues in mind, we wish to reconsider the statistical methodology for inferring the R -ratio from noisy data. As we shall discuss in Sect. 2, the existing approaches use carefully constructed but ad hoc techniques and closed-source software, and consider uncertainties in a frequentist framework. The data-driven approach, though, is connected to common problems in data-science and statistics: modeling an unknown function (here the R -ratio) and managing the risks of under- and over-fitting. In Sect. 3, we describe how we tackle these issues using Gaussian processes – flexible non-parametric statistical models – and marginalization of the model’s hyperparameters.² This allows coherent uncertainty quantification and regularizes the wiggliness of the

R -ratio, which helps prevent the model from over-fitting the noisy data. Our algorithm is implemented in our public `king-pin` package documented in a separate paper [28]. We focus on modeling choices and developing a tool for principled modeling of the R -ratio; our estimates are supplementary to existing ones and we don’t attempt to match previous comprehensive estimates in all respects. We don’t anticipate dramatic differences with respect to previous findings; however, careful modeling of the R -ratio is important because $\mathcal{O}(1\%)$ changes in the HVP contribution or finding that the uncertainty was underestimated could resolve tension with the experimental measurements and lattice predictions. We present predictions from our model for a_{μ}^{HVP} and $\Delta\alpha_{\text{had}}$ in Sect. 4. Finally, we conclude in Sect. 5.

2 Existing data-driven methods

We now briefly review two data-driven methods for calculating a_{μ}^{HVP} . First, the DHMZ approach [29–31], which employs `HVPTools`, a private software package that combines and integrates cross-section data from $e^+e^- \rightarrow$ hadrons. For each experiment, second-order polynomial interpolation is used between adjacent measurements to discretize the results into small bins (of around 1 MeV) for later averaging and numerical integration. The HVP contributions are estimated in a frequentist framework. To ensure that uncertainties are propagated consistently, pseudo-experiments are generated and closure tests with known distributions are performed to validate the data combination and integration. If the results from different experiments are locally inconsistent, the uncertainty of the combination is readjusted according to the local χ^2 value following the well-known PDG approach [32].

The second method is the KNT approach [4, 33, 34], which performs a data-driven compilation of hadronic R -ratio data to calculate the HVP contribution. It first selects the data to be used and then bins the data using a clustering procedure to avoid over-fitting. The clustering procedure determines the optimal binning of data for all channels into a set of clusters based on the available local data density. The optimal clustering criteria are shown in Ref. [4]. As of Ref. [33], the KNT compilation uses an iterated χ^2 fit to achieve the actual combination. This new method ensures that the covariance matrix is re-initialized at each iteration. The motivation of this procedure is to avoid bias. The fit results in the mean R -ratio for each cluster and a full covariance matrix containing all correlated and uncorrelated uncertainties. Combined with trapezoidal integration, these are used to determine channel-by-channel contributions to a_{μ}^{HVP} .

The DHMZ and KNT approaches are both data-driven methods that estimate a_{μ}^{HVP} in a frequentist framework, using privately curated databases of measurements, and in-house

¹ The superscript in σ^0 denotes the bare cross section for e^+e^- annihilation to hadrons, which is defined as the measured cross section that has been corrected for electron-vertex loop contributions, initial state radiation (ISR) and vacuum polarization (VP) effects in the photon propagator.

² Gaussian processes were previously used to smear measurements of the R -ratio [26, 27] to facilitate a comparison with lattice predictions, as energy-smearing predictions are obtainable from lattice QCD. They were not, however, used to model the R -ratio itself.

custom codes and techniques to avoid over-fitting. The differences between the two methods are not only evident in their distinct compilation targets – the DHMZ approach combines and integrates cross-section data from $e^+e^- \rightarrow$ hadrons, while the KNT approach performs a data-driven compilation of the hadronic R -ratio. Furthermore, discernible disparities emerge in their respective data handling procedures, encompassing data selection, data combination, and the propagation of uncertainties. Each method has its own strengths and limitations. While DHMZ and KNT approaches do not exhibit significant differences in estimating the central value of a_μ^{HVP} , there are significant disparities in the resulting uncertainties and the shapes of the combined spectra.

3 Treed Gaussian process

3.1 Gaussian processes

In our data-driven approach, we model the unknown R -ratio with Gaussian processes (GPs; [35,36]). A GP generalizes the Gaussian distribution. Roughly speaking, whereas a Gaussian describes the distribution of a scalar and a multivariate Gaussian describes the distribution of a vector, a GP describes the distribution of a function – an infinite collection of variables $f(x)$ indexed by a location x . Any subset of the random variables are correlated through a multivariate Gaussian. The degree of correlation between $f(x)$ and $f(x')$ governs the smoothness of $f(x)$ and is set by a choice of kernel function, $k(x, x')$.

Just as a GP generalizes a Gaussian distribution of scalars or vectors to a distribution of functions, it allows us to generalize inference over unknown scalars or vectors to inference over unknown functions. Suppose we wish to learn an unknown function. Because a GP describes the distribution of a function, it can be used as the prior for the unknown function in a Bayesian setting. This prior distribution can be updated through Bayes' rule by any noisy measurements or exact calculations of the values of f at particular locations x . In this paper we will update a GP for the R -ratio by the noisy measurements of the R -ratio. We use `celerite2` [37] for ordinary GP computations.

The kernel function is usually stationary, that is, depends only on the Euclidean distance between locations,

$$k(x, x') = k(|x - x'|). \quad (8)$$

Once a particular form of stationary kernel has been chosen, a GP can be controlled by three hyperparameters: a constant mean μ ,

$$\mathbb{E}[f(x)] = \mu, \quad (9)$$

and a scale σ and length ℓ that govern the covariance,

$$\text{Cov}[f(x), f(x')] = \sigma^2 k\left(\frac{|x - x'|}{\ell}\right). \quad (10)$$

The scale controls the size of wiggles in the function predicted by the GP. The length determines the length scale over which correlation decays and hence the number of wiggles in an interval. For Gaussian kernels, by Rice's formula [38] the expected number of wiggles per unit distance scales as $1/\ell$. These three hyperparameters can substantially affect how well a GP models an unknown function. In a fully Bayesian framework, the hyperparameters are marginalized. This automatically weights choices of hyperparameter by how well they model the data and alleviates overfitting. The wiggles is regularized and the fit needn't pass through every data point.

3.2 Treed Gaussian processes

The GPs described thus far are stationary – they model all regions of input space identically. To allow for non-stationary structure in the R -ratio, we use a treed-GP (TGP; [39,40]). This is necessary as we know that the R -ratio contains narrow features such as resonances. In a TGP, the input space is partitioned using a binary tree. The predictions in each partition are governed by a different GP with independent hyperparameters. The number and locations of partitions are modeled using the so-called CGM prior [41].

The difference in predictions between a GP and our TGP is illustrated in Fig. 1. In this illustration we consider evenly spaced noisy measurements of a function that contains a step. The GP (left) models the data poorly, as to accommodate the sudden step, the covariance between input locations must be weak which results a wiggly fit to the straight-line sections. The TGP (right) automatically partitions the input space allowing it to model the distinct straight-line sections separately. At the jumps before and after the step, the TGP predicts the function with substantial uncertainty. This is satisfying since the data points change dramatically across those regions of input space and do not indicate what might happen inside them.

TGPs build on ideas such as CART [41], treed models generally [42] and partitioning [43], and are similar to piecewise GPs [44] and a recent proposal in machine learning [45]. Alternative approaches to non-stationarity include non-stationary kernel functions [46], Deep GPs [47,48] where non-stationarity is modeled through warping, and hierarchical models of GPs [49,50]. There is valuable discussion and comparison of these approaches in Refs. [51,52] and this remains an active area. As well as addressing non-stationarity in unknown functions, these approaches address heteroscedastic noise in our measurements.

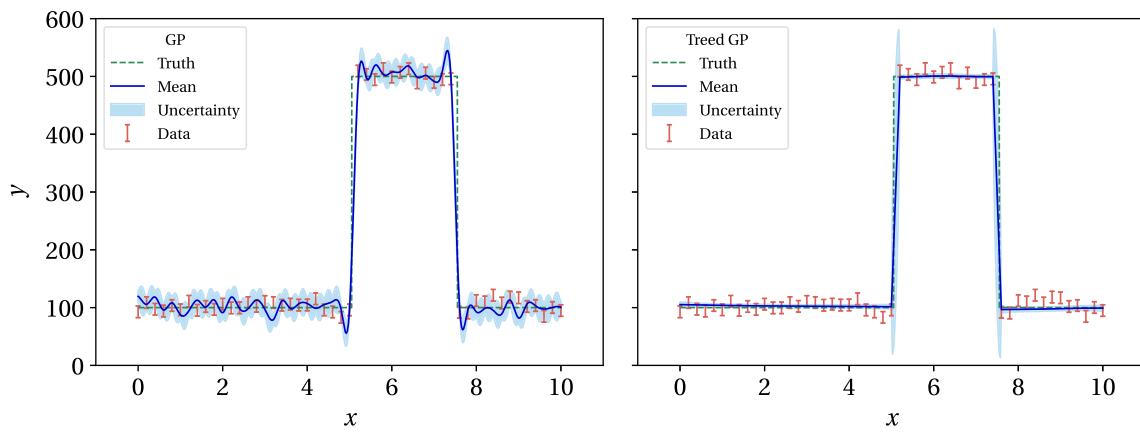


Fig. 1 The GP (left) predicts a wiggly fit to the straight line sections due to the step. The TGP (right) automatically addresses the issue by partitioning the input space

Our approach is fully Bayesian – we marginalize the GP hyperparameters and tree structure. This decreases the risk of over- or under-fitting the noisy data and smooths the partitions between GPs. We perform marginalization numerically using reversible jump Markov Chain Monte Carlo (RJMCMC; [53]). For reviews see Refs. [54–56]). This is a generalization of MCMC that works on parameter spaces that don’t have a fixed dimension – this is vital because the number of GPs and thus the total number of hyperparameters isn’t fixed. Navigating the tree structure requires special RJMCMC proposals – such as growing, pruning and rotating the tree – that are described in Ref. [39].

3.3 Integration

The idea of modeling integrals through GPs was originally known by Bayes–Hermite quadrature [57], and later discussed under the names of Bayesian Monte Carlo (BMC; [58]) and Bayesian quadrature or cubature [59,60]; see Ref. [61] for a review. Suppose we wish to compute an integral of the form,

$$I = \int C(x)f(x) dx \tag{11}$$

where $f(x)$ is the estimated function and $C(x)$ is a known function. BMC provides an epistemic meaning to errors in quadrature estimates of these integrals, such as

$$I \approx \sum_i C(x_i)f(x_i)\Delta x_i, \tag{12}$$

because we may make inferences on I through our statistical model for $f(x)$. In cases in which the function $C(x)$ and choice of kernel lead to intractable computations, there is an additional discretization error in BMC inferences as the GP predictions are evaluated on a finitely-spaced grid. This is known as approximate Bayesian cubature [61]. This additional error may be neglected when the integrand is approx-

imately linear between prediction points. We will use a TGP to model an integrand. Although trees have been proposed in BMC [62], they haven’t previously been directly combined with GPs in this way.

3.4 Sequential design

After completing inference of an unknown function with the data at hand, one may wish to know what data to collect next. This problem is known as sequential design or active learning. Broadly speaking, this is a challenging question and greedy approaches that make optimal choices one step at a time are easier to implement. We thus consider a variant of active learning Mackay (ALM; [63,64]).

Following the approach in Ref. [65], we consider the location that contributes most to the uncertainty in the a_μ^{HVP} and $\Delta\alpha_{\text{had}}$ integrals to be an optimal location at which to perform more measurements. For an integral of the form eq. (11), we compute

$$x_{\text{ALM}} = \arg \max_x \left[\int C(x) \text{Cov}[f(x), f(y)] C(y) dy \right]. \tag{13}$$

See Refs. [40,66] for further discussion.

4 Results

4.1 Data selection

We investigated the public dataset from the Particle Data Group (PDG; [32,67]). See also Ref. [68] for further details, which primarily comprises data on the inclusive R -ratio, asymmetric statistical errors, and point-to-point systematic errors from electron-positron annihilation to hadrons at different CM energies. Certain CM energies may have multiple

point-to-point systematic uncertainties stemming from different sources. We symmetrized errors and combined systematic (τ) and statistical errors (σ) in quadrature:

$$\tau^2 = \sum_i \tau_i^2 \quad \text{where} \quad \tau_i = \frac{1}{2}(\tau_{\text{up}} + |\tau_{\text{down}}|), \quad (14)$$

$$\sigma_{\text{total}} = \sqrt{\sigma^2 + \tau^2} \quad \text{where} \quad \sigma = \frac{1}{2}(\sigma_{\text{up}} + |\sigma_{\text{down}}|). \quad (15)$$

We selected 859 data points inside the CM energy interval 0.3–1.937 GeV. This interval was selected to facilitate a comparison with Ref. [33]. The maximum $\sqrt{s} = 1.937$ GeV was chosen as it is the point at which summing exclusive R -ratio data becomes unfeasible and perturbative QCD may be reliable. The minimum $\sqrt{s} = 0.3$ GeV was the minimum energy in the public PDG dataset.

To model this data set, we utilized a treed Gaussian process (TGP), as described in Sect. 3. Besides selecting the data, we must specify the locations at which we want to predict the R -ratio. In our study, we predicted at every input location and at two uniformly spaced locations between every pair of consecutive input locations.

4.2 Computational methods and modelling choices

We model the R -ratio by a TGP in which the input space is divided into partitions using a binary tree. Each partition in our TGP is governed by a mean, μ , and a Matérn-3/2 kernel with independent scale, σ , and length, ℓ , hyperparameters. We use a uniform prior between 0 and 150 for the mean, a uniform prior between 0 and 500 for the scale, and a uniform prior between 0–5 GeV for the length. These choices were motivated by the maximum measured R -ratio and the CM interval 0.3–1.937 GeV under consideration. Following Ref. [39], the structure of the tree itself is controlled by a CGM prior with hyperparameters $\alpha = 0.5$ and $\beta = 2$; see Ref. [41] for explanation of these parameters. These choices favor smaller and more balanced trees.

We marginalize the tree structure and hyperparameters using RJ-MCMC. To improve computational efficiency, we thin the chains by a factor of four and only compute predictions for the states in the thinned chains. This reduces the computational time but only slightly reduces the effective sample size as the states in the unthinned chain are strongly correlated. We run RJ-MCMC for 300,000 steps but discard 5000 burn-in steps to minimize bias from the beginning of the chain. For computational efficiency and following a multistart heuristic, we run 10 chains in parallel and combine them.

4.3 Predictions

The predictions from our TGP model for the R -ratio are shown in Fig. 2 as a mean and an error band. The mean predictions pass smoothly around the data points without any undue fluctuations near the data points that are characteristic of over-fitting. The ρ - ω and ϕ resonances are typically fitted by their own tree partitions with separate hyperparameters. They aren't forced to be as smooth as the rest of the spectrum and appear well-fitted. Our model predictions are noticeably more uncertain in regions with fewer or noisy measurements. We identify no anomalous features and tentatively conclude that the RJ-MCMC marginalization adequately converged. We ran standard MCMC diagnostics on the mean of the R -ratio using ArviZ [69]; finding $n_{\text{eff}} \simeq 600$ bulk effective samples and Gelman–Rubin diagnostic 1.01 [70]. There are typically five or six partitions, as the two peaks and the three flatter regions are modeled separately as shown in Fig. 3. In Fig. 4 we show the result when using an ordinary GP. To accommodate the narrow peaks in the measured R -ratio, the GP model permits substantial wiggles between data points, especially where the data points are sparse.

The TGP model outputs the mean, $\mathbb{E}[R_i]$, and covariance, $\text{Cov}[R_i, R_j]$, of the R -ratio at the prediction locations \sqrt{s}_i . The mean function represents the expected or average output value for a given input value, while the covariance function represents the covariance between predictions at different CM energies. As the RJ-MCMC can be computationally expensive and time-consuming, we saved these results to disk and made them publicly available [71]. We used the mean and covariance predictions for R in combination with the dispersion integrals to predict contributions to a_μ^{HVP} and $\Delta\alpha_{\text{had}}$ from the CM energy interval of 0.3–1.937 GeV. As a_μ^{HVP} and $\Delta\alpha_{\text{had}}$ are linear functions of R , we propagate $\mathbb{E}[R_i]$ and $\text{Cov}[R_i, R_j]$ to obtain predictions. In all subsequent integration processes, we employ the trapezoidal rule [72], and in subsequent formulae \sqrt{s} denotes the locations of our TGP predictions rather than the locations of the measurements.

The calculation of a_μ^{HVP} is based on Eq. (4). However, since the independent variable in this case is the CM energy, a simple deformation of Eq. (4) is necessary,

$$a_\mu^{\text{HVP}} = \frac{2\alpha^2}{3\pi^2} \int_{m_\pi^2}^\infty \frac{K(s)R(\sqrt{s})}{\sqrt{s}} d\sqrt{s}. \quad (16)$$

Then we calculate the value of a_μ^{HVP} through numerical quadrature,

$$a_\mu^{\text{HVP}} = \sum_i C_i^{\text{HVP}} R_i, \quad (17)$$

where we defined

$$C_i^{\text{HVP}} \equiv \frac{2\alpha^2}{3\pi^2} \frac{K(\sqrt{s}_i)}{\sqrt{s}_i} w_i, \quad (18)$$

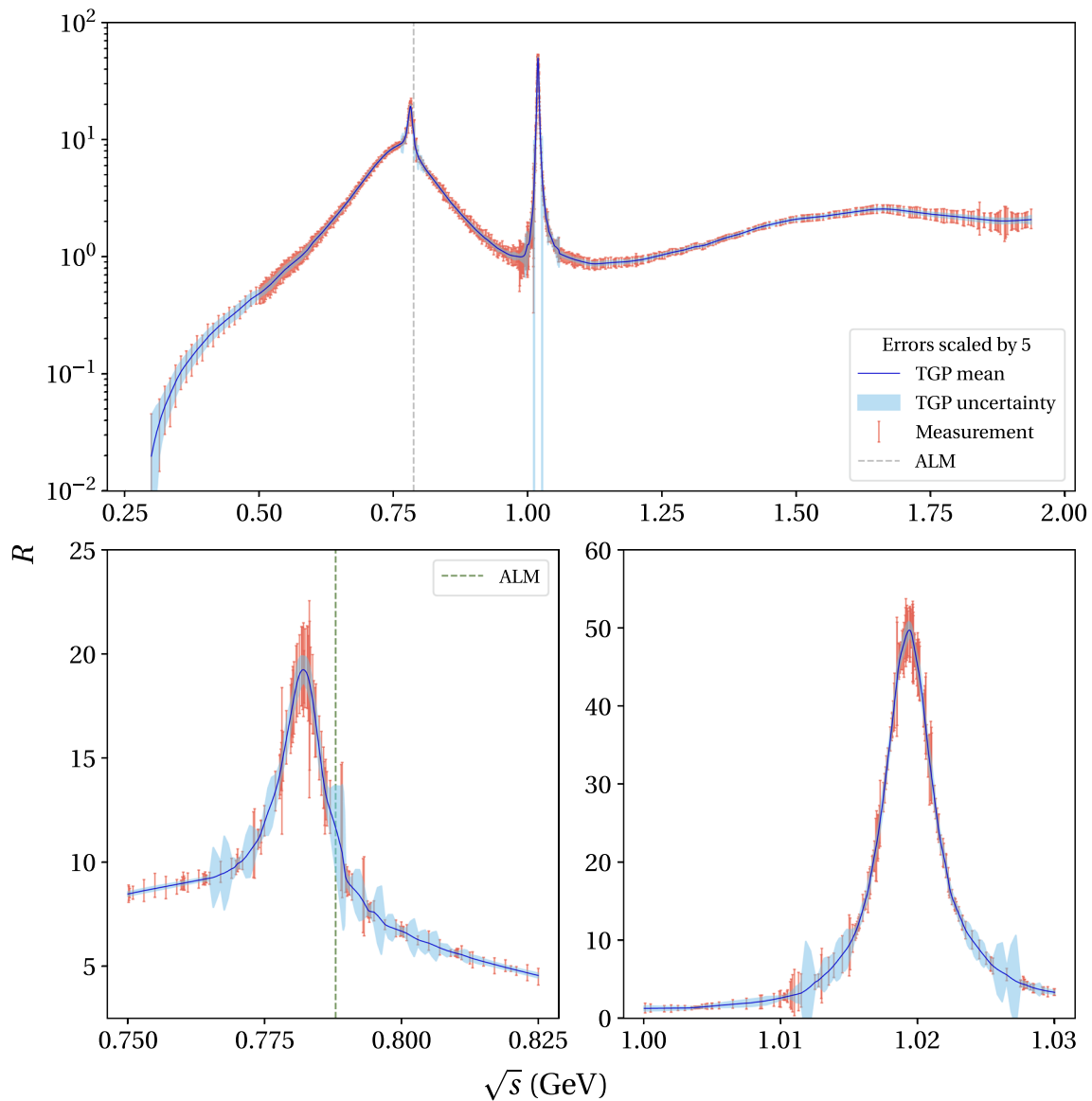


Fig. 2 Predicted R -ratio from the TGP model. The experimental errors and uncertainty in the TGP predictions are scaled by 5 and the $\rho - \omega$ and ϕ resonances are plotted separately for visibility

where w_i are the quadrature weights. We use the trapezoid rule such that

$$w_i = \frac{1}{2} \begin{cases} \sqrt{s_2} - \sqrt{s_1} & i = 1 \\ \sqrt{s_N} - \sqrt{s_{N-1}} & i = N \\ \sqrt{s_{i+1}} - \sqrt{s_{i-1}} & \text{otherwise.} \end{cases} \quad (19)$$

From Eq. (17) and by linearity, the mean can be found through

$$\mathbb{E}[a_\mu^{\text{HVP}}] = \sum_i C_i^{\text{HVP}} \mathbb{E}[R_i], \quad (20)$$

and from the covariance matrix for predictions of the R -ratio, the uncertainty in our prediction of a_μ^{HVP} can be calculated

using

$$\begin{aligned} \text{Var}[a_\mu^{\text{HVP}}] &= \text{Var}\left[\sum_i C_i^{\text{HVP}} R_i\right] \\ &= \sum_{i,j=1}^n C_i^{\text{HVP}} C_j^{\text{HVP}} \text{Cov}[R_i, R_j]. \end{aligned} \quad (21)$$

We compute $\Delta\alpha_{\text{had}}$ similarly using,

$$\Delta\alpha_{\text{had}} = \frac{2\alpha M_Z^2}{3\pi} \int_{m_\pi^2}^\infty \frac{R(\sqrt{s})}{\sqrt{s}(M_Z^2 - (\sqrt{s})^2)} d\sqrt{s}. \quad (22)$$

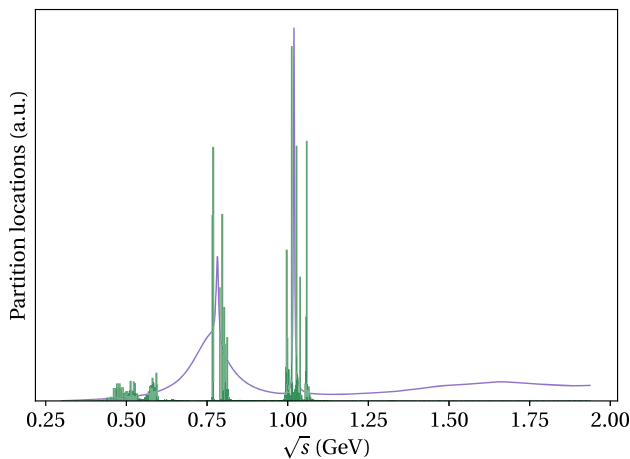


Fig. 3 Histogram of locations of partition edges in the TGP model. The mean prediction for the R -ratio is shown for reference (blue)

We use Eqs. (20) and (21) though with coefficients,

$$C_i^{\text{had}} = \frac{2\alpha M_Z^2}{3\pi} \frac{w_i}{\sqrt{s_i} [M_Z^2 - s_i]} \tag{23}$$

Because our calculation is performed at CM energies from 0.3 to 1.937 GeV, the principal-value prescription does not need to be considered.

For the sake of comparison and to verify parts of our tool-chain, we calculate a_μ^{HVP} and $\Delta\alpha_{\text{had}}$ naively without utilizing a TGP. We consider a naive model that at the locations of the measurements of the R -ratio predicts

$$R_i = \hat{R}_i \pm \sigma_i \tag{24}$$

where \hat{R}_i are the central values and σ_i are the errors of the measurements. In this naive model there is no covariance between predictions, that is, $\text{Cov}[R_i, R_j] = 0$ for $i \neq j$. Equations (20) and (21) apply to this simple case, although it should be noted that in this case \sqrt{s} are a series of data points, whereas in the TGP \sqrt{s} are the chosen prediction locations.

The results of the above calculations are summarized in Table 1. We show the predictions from KNT18 [33] and KNT19 [34] for comparison, which are found by summing data-based exclusive channels in Tables 2 and 1, respectively, and combing errors in quadrature.³ We see that the TGP prediction for a_μ^{HVP} is smaller than predictions from the naive model, KNT18 [33] and KNT19 [34]. This would make tension between data-driven estimates and lattice QCD and the experimental measurements worse. The uncertainties in our TGP predictions are nearly identical to those from the naive model – we explain this similarity in uncertainties in appendix A – though substantially smaller than those from KNT18 [33] and KNT19 [34]. We don’t anticipate

³ Compared to KNT18 [33], several channels were updated and two new channels measured by CMD-3 [14] were included in KNT19 [34].

that the smaller TGP uncertainties are a consequence of the TGP model itself; rather, KNT18 [33] and KNT19 [34] are based on a different dataset and treatment of systematics. For example, they include uncertainties from vacuum polarization (VP) effects and final-state radiation (FSR) that we omit. We thus find no clear evidence of mismodelling or that our more careful modeling can shed light on the tension between data-driven estimates, lattice estimates and experiments. It is possible, however, that for an identical dataset to that in KNT18 [33] and KNT19 [34], the TGP predictions could be greater than KNT18 [33] and KNT19 [34] – the impact of reducing overfitting with a TGP could work in the opposite direction in that dataset.

4.4 Sequential design

We may use our TGP result to identify locations that contribute most to the uncertainty in the a_μ^{HVP} and $\Delta\alpha_{\text{had}}$ predictions and where future measurements would be most beneficial. For both a_μ^{HVP} and $\Delta\alpha_{\text{had}}$, the ALM estimate from Eq. (13) yields

$$\sqrt{s}_{\text{ALM}} = 0.788 \text{ GeV}. \tag{25}$$

This lies near noisy measurements after the $\rho - \omega$ resonance; see Fig. 2. Besides lying close to noisy measurements, the uncertainty at this location is substantial because it is a boundary between partitions of the TGP – the behavior of the function changes abruptly here and so is hard to predict.

4.5 Correlation

Lastly, let us consider the relationship between the predictions for a_μ^{HVP} and $\Delta\alpha_{\text{had}}$. From Eqs. (4) and (5), we observe that the dispersion integral formulas used to calculate a_μ^{HVP} and $\Delta\alpha_{\text{had}}$ both involve the R -ratio. To quantify this relationship, covariance can be used to measure the correlation between two variables. The sign of covariance indicates whether the trends between the two variables are consistent. The correlation coefficient is usually utilized to reflect the strength of the correlation between two variables. Thus, to gain more understanding about the relationship between a_μ^{HVP} and $\Delta\alpha_{\text{had}}$, we computed their covariance and correlation,

$$\text{Cov}[a_\mu^{\text{HVP}}, \Delta\alpha_{\text{had}}] = \sum_{i,j=1}^n C_i^{\text{HVP}} C_j^{\text{had}} \text{Cov}[R_i, R_j] \tag{26}$$

$$\rho[a_\mu^{\text{HVP}}, \Delta\alpha_{\text{had}}] = \frac{\text{Cov}[a_\mu^{\text{HVP}}, \Delta\alpha_{\text{had}}]}{\sqrt{\text{Var}[a_\mu^{\text{HVP}}] \text{Var}[\Delta\alpha_{\text{had}}]}} \tag{27}$$

where the (co)variances were computed under the TGP as described. As anticipated, we obtained a positive correlation between the two. Specifically, when $\Delta\alpha_{\text{had}}$ increases,

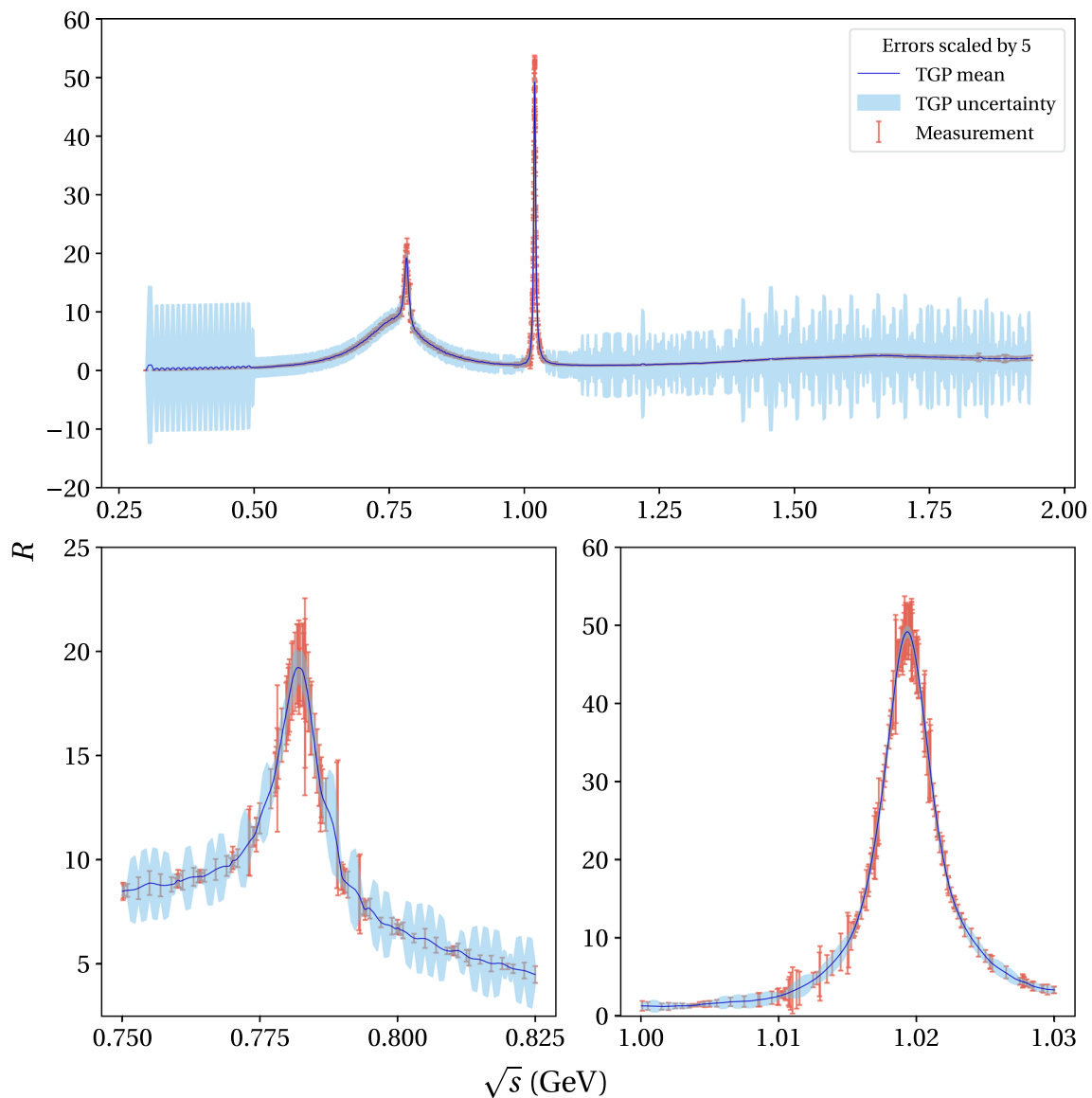


Fig. 4 Similar to Fig. 2, though showing results from an ordinary GP

Table 1 Contributions to a_μ^{HVP} and $\Delta\alpha_{\text{had}}$ in the CM energy range of 0.3–1.937 GeV from our TGP model, a naive model, KNT18 [33] and KNT19 [34]

	Treed GP model	Naive model	KNT18 [33]	KNT19 [34]
$a_\mu^{\text{HVP}} \times 10^{10}$	636.34 ± 0.83	637.73 ± 0.84	638.59 ± 2.35	638.25 ± 2.32
$\Delta\alpha_{\text{had}} \times 10^4$	57.28 ± 0.05	57.41 ± 0.05	57.83 ± 0.29	57.88 ± 0.29

the value of a_μ^{HVP} also increases, and vice versa. The calculated correlation coefficient was $\rho \simeq 0.8$, which is close to 1, quantifying the strong correlation between a_μ^{HVP} and $\Delta\alpha_{\text{had}}$.

5 Discussion and conclusions

The BNL and FNAL measurements of the anomalous magnetic moment of the muon disagree with the Standard Model

(SM) prediction by more than 4σ . This has led to renewed scrutiny of new physics explanations and the SM prediction. With that as motivation, we extracted the hadronic vacuum polarization (HVP) contributions, a_μ^{HVP} , from electron cross-section data using a treed Gaussian process (TGP) to model the unknown R -ratio as a function of CM energy. This is a principled and general method from data-science, that allows complete uncertainty quantification and automatically balances over- and under-fitting to noisy data.

The challenges in the data-driven approach are common in data-science. A competitive estimate of a_μ^{HVP} , however, requires domain-specific expertise, careful curation of measurements, and careful consideration of systematic errors and their correlation. This should be developed over time in collaboration with domain experts. Thus our work should be seen as preliminary and serves to explore an alternative statistical methodology based on more general principles and develop an associated toolchain. We used a dataset available from the PDG, though as noted as early as 2003 in Ref. [68], a more complete, documented and standardized database of measurements would allow further scrutiny of data-driven estimates of HVP.

Our analysis used about $n \approx 1000$ data points. The linear algebra operations in GP computations scale as $\mathcal{O}(n^3)$. There are computational approaches and approximations to overcome this scaling (see e.g., Refs. [73–77]); nevertheless, working with more complete datasets could be challenging. On the other hand, splitting data channel by channel could help the situation. For a competitive estimate, we would require careful treatment of correlated systematic uncertainties. The approach started here – carefully building an appropriate statistical model – naturally allows us to model systematic uncertainties. For example, through nuisance parameters for scale uncertainties or sophisticated noise models for correlated noise (see e.g., Ref. [78]). The statistical model could include, for example, a hierarchical model of systematic uncertainties accounting for “errors on errors.”

The prediction for a_μ^{HVP} from our TGP model is slightly smaller than existing data-driven estimates. Thus, more principled modeling of the R -ratio in fact increases tension between the SM prediction and measurements for $g - 2$. On the other hand, because the kernel functions were slowly-varying, the TGP model predicted a_μ^{HVP} with a similar uncertainty to that obtained in naive approach. This can be understood from the trade-off between variance and covariance in predictions of the R -ratio at different CM energies. Looking forward, by the ALM criteria, the best CM energy for future measurements was $\sqrt{s} \simeq 0.788 \text{ GeV}$ for both a_μ^{HVP} and $\Delta\alpha_{\text{had}}$, as it lies close to particularly noisy measurements of the R -ratio. In conclusion, we developed a statistical model for the R -ratio, based on general principles and publicly available toolchains. We found no indication that mismodeling the R -ratio could be responsible for tension with measurements or lattice predictions. We hope, however, that this work serves as a starting point for further scrutiny, principled modeling and development of associated public tools.

Acknowledgements AF was supported by RDF-22-02-079. We thank Peter Athron for comments and feedback.

Data Availability Statement This manuscript has associated data in a data repository. [Authors’ comment: The TGP algorithm is imple-

mented in our public Python package `kingpin` [28]. The dataset and codes, which use the `kingpin` package, for this paper are available online; see Ref. [71].]

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/>.

Funded by SCOAP³. SCOAP³ supports the goals of the International Year of Basic Sciences for Sustainable Development.

Appendix A: Similarity between TGP and naive model uncertainties

Here we explain why the uncertainties in the predictions for a_μ^{HVP} and $\Delta\alpha_{\text{had}}$ calculated under our TGP and in the naive method are approximately equal,

$$\sigma_{\text{TGP}}^2 = \sum_{i,j=1}^n C_i C_j \text{Cov}[R_i, R_j] \tag{28}$$

$$\simeq \sigma_{\text{naive}}^2 = \sum_{i=1}^n C_i^2 \sigma_i^2, \tag{29}$$

despite quite different predictions for the R -ratio. We anticipated a reduction in error in our TGP as we applied prior information about correlations between prediction points. There are two effects of introducing correlation: first, as anticipated a reduction in the variance at any prediction point, that is, $\text{Var}[R_i] \ll \sigma_i^2$. Second, an increase in the correlation between prediction points, $\text{Cov}[R_i, R_j] > 0$. The uncertainty in our TGP predictions includes terms containing variance and covariance,

$$\begin{aligned} \sigma_{\text{TGP}}^2 &= \sum_{i,j=1}^n C_i C_j \text{Cov}[R_i, R_j] \\ &= \sum_{i=1}^n C_i^2 \text{Var}[R_i] + \sum_{i \neq j} C_i C_j \text{Cov}[R_i, R_j] \end{aligned} \tag{30}$$

In practice we find that the decrease in the variance is almost exactly canceled by the increase in the covariance.

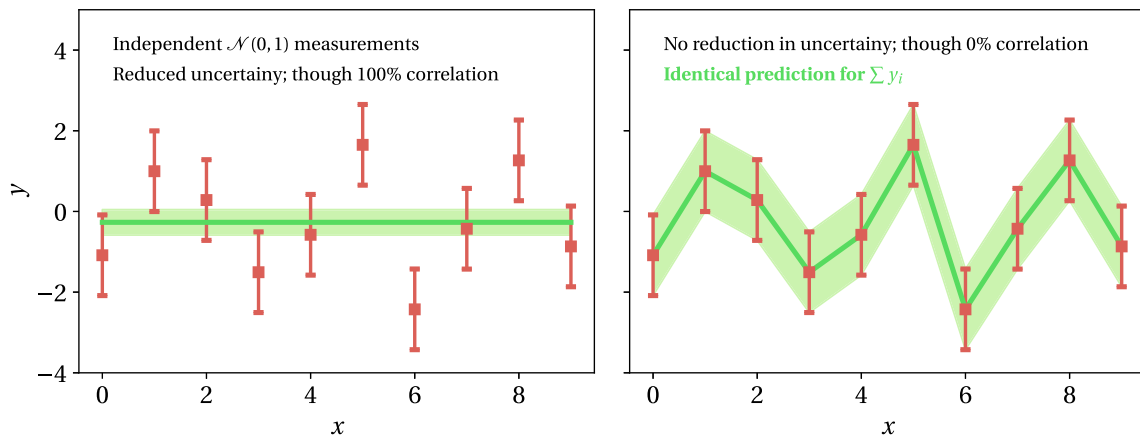


Fig. 5 A horizontal line fitted to noisy data (left) and a naive model that passes through every data point (right). Despite making quite different predictions for y , they make identical predictions for $\sum y_i$

To understand this effect in the simplest way, consider fitting a horizontal line ($m = 0$) with an unknown intercept (c) through data points with errors σ as shown in Fig. 5. The uncertainty at any prediction point equals the uncertainty on the intercept,

$$\text{Var}[y_i] = \frac{\sigma^2}{n} \tag{31}$$

This is reduced from σ^2 by the 100% correlation between the predictions at the n data points. Now consider the uncertainty on the sum,

$$\begin{aligned} \text{Var}\left[\sum_{i=1}^n y_i\right] &= \sum_{i,j=1}^n \text{Cov}[y_i, y_j] \\ &= \sum_{i=1}^n \text{Var}[y_i] + \sum_{i \neq j} \text{Cov}[y_i, y_j] \end{aligned} \tag{32}$$

$$= n \frac{\sigma^2}{n} + (n^2 - n) \frac{\sigma^2}{n} = n\sigma^2 \tag{33}$$

This is identical to the uncertainty from our naive model with no correlations that fits $y = \hat{y} \pm \sigma$ because the increased covariance cancels the decreased variance. Although we reduced the uncertainty at any prediction point, that reduction was offset by the covariance between predictions.

Let us create an example closer to our TGP model and demonstrate an identical effect. In a GP with fixed hyperparameters for measurements with uniform noise σ , the covariance between prediction points X^* can be expressed as

$$\begin{aligned} \text{Cov}[y(X^*), y(X^*)] &= K(X^*, X^*) - K(X^*, X)^T \\ &\quad \left[K(X, X) + \sigma^2 \right]^{-1} K(X^*, X) \end{aligned} \tag{34}$$

where X are the training points, K is the choice of kernel function, and σ^2 is the noise. This expression is somewhat

intractable due to the inverse matrix. Thus we consider $X = X^*$ and a simplified covariance,

$$K_{ij} = a + b\delta_{ij}. \tag{35}$$

With this choice, one can compute $\text{Cov}[y(X^*), y(X^*)]$ and the sum over its elements analytically. The covariance may be written as

$$\text{Cov}[y(X^*), y(X^*)] = A \begin{pmatrix} D & a\sigma^2 & \dots & a\sigma^2 \\ a\sigma^2 & D & \dots & a\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ a\sigma^2 & a\sigma^2 & \dots & D \end{pmatrix} \tag{36}$$

where

$$A \equiv \frac{\sigma^2}{(b + \sigma^2)(na + b + \sigma^2)} \tag{37}$$

$$D \equiv b(na + b) + (a + b)\sigma^2. \tag{38}$$

Summing the elements results in,

$$\sum \text{Cov}[y(X^*), y(X^*)] = \frac{n(na + b)\sigma^2}{na + b + \sigma^2} \tag{39}$$

where n represents the size of X^* . By performing a Taylor expansion about $\sigma = 0$, we discover

$$\sum \text{Cov}[y(X^*), y(X^*)] = n\sigma^2 + \mathcal{O}(\sigma^4). \tag{40}$$

Although the structure Eq. (35) isn't particularly realistic, our result holds for any a , including 0% and 100% correlation. When $C_i \approx C \equiv \text{const.}$, the uncertainties from the TGP and naive method are both around $C\sqrt{n}\sigma$. The decrease in variance and increase in covariance cancel in the TGP uncertainties. For the case in which the prediction locations are denser than the input locations, this result may hold if the TGP predictions don't change substantially between input locations.

References

1. The Muon $g - 2$ Collaboration, Final report of the Muon E821 anomalous magnetic moment measurement at BNL. *Phys. Rev. D* **73**, 072003 (2006). <https://doi.org/10.1103/PhysRevD.73.072003>. [arXiv:hep-ex/0602035](https://arxiv.org/abs/hep-ex/0602035)
2. The Muon $g - 2$ Collaboration, Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.* **126**, 141801 (2021). <https://doi.org/10.1103/PhysRevLett.126.141801>. [arXiv:2104.03281](https://arxiv.org/abs/2104.03281)
3. P. Athron, C. Balázs, D.H.J. Jacob, W. Kotlarski, D. Stöckinger, H. Stöckinger-Kim, New physics explanations of a_μ in light of the FNAL muon $g - 2$ measurement. *JHEP* **09**, 080 (2021). [https://doi.org/10.1007/JHEP09\(2021\)080](https://doi.org/10.1007/JHEP09(2021)080). [arXiv:2104.03691](https://arxiv.org/abs/2104.03691)
4. T. Aoyama et al., The anomalous magnetic moment of the muon in the Standard Model. *Phys. Rep.* **887**, 1 (2020). <https://doi.org/10.1016/j.physrep.2020.07.006>. [arXiv:2006.04822](https://arxiv.org/abs/2006.04822)
5. P. Athron, A. Fowlie, C.-T. Lu, L. Wu, Y. Wu, B. Zhu, Hadronic uncertainties versus new physics for the W boson mass and Muon $g - 2$ anomalies. *Nat. Commun.* **14**, 659 (2023). <https://doi.org/10.1038/s41467-023-36366-7>. [arXiv:2204.03996](https://arxiv.org/abs/2204.03996)
6. T. Teubner, K. Hagiwara, R. Liao, A.D. Martin, D. Nomura, Update of $g - 2$ of the muon and delta alpha. *Chin. Phys. C* **34**, 728 (2010). <https://doi.org/10.1088/1674-1137/34/6/019>. [arXiv:1001.5401](https://arxiv.org/abs/1001.5401)
7. The KLOE-2 Collaboration, Combination of KLOE $\sigma(e^+e^- \rightarrow \pi^+\pi^-\gamma(\gamma))$ measurements and determination of $a_\mu^{\pi^+\pi^-}$ in the energy range $0.10 < s < 0.95 \text{ GeV}^2$. *JHEP* **03**, 173 (2018). [https://doi.org/10.1007/JHEP03\(2018\)173](https://doi.org/10.1007/JHEP03(2018)173). [arXiv:1711.03085](https://arxiv.org/abs/1711.03085)
8. The BaBar Collaboration, Study of the process $e^+e^- \rightarrow \pi^+\pi^-\eta$ using initial state radiation. *Phys. Rev. D* **97**, 052007 (2018). <https://doi.org/10.1103/PhysRevD.97.052007>. [arXiv:1801.02960](https://arxiv.org/abs/1801.02960)
9. The BaBar Collaboration, The $e^+e^- \rightarrow 2(\pi^+\pi^-)\pi^0$, $2(\pi^+\pi^-)\eta$, $K^+K^-\pi^+\pi^-\pi^0$ and $K^+K^-\pi^+\pi^-\eta$ cross sections measured with initial-state radiation. *Phys. Rev. D* **76**, 092005 (2007). <https://doi.org/10.1103/PhysRevD.76.092005>. [arXiv:0708.2461](https://arxiv.org/abs/0708.2461). [Erratum: *Phys. Rev. D* **77**, 119902 (2008)]
10. The BaBar Collaboration, Measurement of the $e^+e^- \rightarrow \pi^+\pi^-\pi^0$ cross section using initial-state radiation at BaBar. *Phys. Rev. D* **96**, 092009 (2017). <https://doi.org/10.1103/PhysRevD.96.092009>. [arXiv:1709.01171](https://arxiv.org/abs/1709.01171)
11. The BaBar Collaboration, Study of the reactions $e^+e^- \rightarrow \pi^+\pi^-\pi^0\pi^0\gamma$ and $\pi^+\pi^-\pi^0\pi^0\eta\gamma$ at center-of-mass energies from threshold to 4.35 GeV using initial-state radiation. *Phys. Rev. D* **98**, 112015 (2018). <https://doi.org/10.1103/PhysRevD.98.112015>. [arXiv:1810.11962](https://arxiv.org/abs/1810.11962)
12. The SND Collaboration, Study of the reaction $e^+e^- \rightarrow \pi^0\gamma$ with the SND detector at the VEPP-2M collider. *Phys. Rev. D* **93**, 092001 (2016). <https://doi.org/10.1103/PhysRevD.93.092001>. [arXiv:1601.08061](https://arxiv.org/abs/1601.08061)
13. The CMD-3 Collaboration, Study of the process $e^+e^- \rightarrow \pi^+\pi^-\pi^0\eta$ in the c.m. energy range 1394–2005 MeV with the CMD-3 detector. *Phys. Lett. B* **773**, 150 (2017). <https://doi.org/10.1016/j.physletb.2017.08.019>. [arXiv:1706.06267](https://arxiv.org/abs/1706.06267)
14. The CMD-3 Collaboration, Study of the process $e^+e^- \rightarrow 3(\pi^+\pi^-\pi^0)$ in the C.M. Energy range 1.6–2.0 GeV with the CMD-3 detector. *Phys. Lett. B* **792**, 419 (2019). <https://doi.org/10.1016/j.physletb.2019.04.007>. [arXiv:1902.06449](https://arxiv.org/abs/1902.06449)
15. B.E. Lautrup, E. De Rafael, Calculation of the sixth-order contribution from the fourth-order vacuum polarization to the difference of the anomalous magnetic moments of muon and electron. *Phys. Rev.* **174**, 1835 (1968). <https://doi.org/10.1103/PhysRev.174.1835>
16. S.J. Brodsky, E. De Rafael, Suggested boson-lepton pair couplings and the anomalous magnetic moment of the muon. *Phys. Rev.* **168**, 1620 (1968). <https://doi.org/10.1103/PhysRev.168.1620>
17. H. Wittig, Progress on $(g - 2)_\mu$ from Lattice QCD, in Proceedings of the 2023 Electroweak Session of the 57th Rencontres de Moriond, June, 2023. [arXiv:2306.04165](https://arxiv.org/abs/2306.04165)
18. G. Benton, D. Boito, M. Golterman, A. Keshavarzi, K. Maltman, S. Peris, Data-driven determination of the light-quark connected component of the intermediate-window contribution to the muon $g - 2$. [arXiv:2306.16808](https://arxiv.org/abs/2306.16808)
19. S. Borsanyi et al., Leading hadronic contribution to the muon magnetic moment from lattice QCD. *Nature* **593**, 51 (2021). <https://doi.org/10.1038/s41586-021-03418-1>. [arXiv:2002.12347](https://arxiv.org/abs/2002.12347)
20. M. Cè et al., Window observable for the hadronic vacuum polarization contribution to the muon $g - 2$ from lattice QCD. *Phys. Rev. D* **106**, 114502 (2022). <https://doi.org/10.1103/PhysRevD.106.114502>. [arXiv:2206.06582](https://arxiv.org/abs/2206.06582)
21. The Extended Twisted Mass Collaboration, Lattice calculation of the short and intermediate time-distance hadronic vacuum polarization contributions to the muon magnetic moment using twisted-mass fermions. *Phys. Rev. D* **107**, 074506 (2023). <https://doi.org/10.1103/PhysRevD.107.074506>. [arXiv:2206.15084](https://arxiv.org/abs/2206.15084)
22. T. Blum et al., An update of Euclidean windows of the hadronic vacuum polarization. [arXiv:2301.08696](https://arxiv.org/abs/2301.08696)
23. P. Masjuan, A. Miranda, P. Roig, τ data-driven evaluation of Euclidean windows for the hadronic vacuum polarization. [arXiv:2305.20005](https://arxiv.org/abs/2305.20005)
24. The CMD-3 Collaboration, Measurement of the $e^+e^- \rightarrow \pi^+\pi^-$ cross section from threshold to 1.2 GeV with the CMD-3 detector. [arXiv:2302.08834](https://arxiv.org/abs/2302.08834)
25. The CMD-2 Collaboration, High-statistics measurement of the pion form factor in the rho-meson energy range with the CMD-2 detector. *Phys. Lett. B* **648**, 28 (2007). <https://doi.org/10.1016/j.physletb.2007.01.073>. [arXiv:hep-ex/0610021](https://arxiv.org/abs/hep-ex/0610021)
26. M. Hansen, A. Lupo, N. Tantalo, Extraction of spectral densities from lattice correlators. *Phys. Rev. D* **99**, 094508 (2019). <https://doi.org/10.1103/PhysRevD.99.094508>. [arXiv:1903.06476](https://arxiv.org/abs/1903.06476)
27. The Extended Twisted Mass Collaboration (ETMC) Collaboration, Probing the energy-smeared R ratio using lattice QCD. *Phys. Rev. Lett.* **130**, 241901 (2023). <https://doi.org/10.1103/PhysRevLett.130.241901>. [arXiv:2212.08467](https://arxiv.org/abs/2212.08467)
28. A. Fowlie, kingpin—treed Gaussian process algorithm (2023). <https://github.com/andrewfowlie/kingpin>
29. M. Davier, A. Hoecker, B. Malaescu, C.Z. Yuan, Z. Zhang, Reevaluation of the hadronic contribution to the muon magnetic anomaly using new $e^+e^- \rightarrow \pi^+\pi^-$ cross section data from BaBar. *Eur. Phys. J. C* **66**, 1 (2010). <https://doi.org/10.1140/epjc/s10052-010-1246-1>. [arXiv:0908.4300](https://arxiv.org/abs/0908.4300)
30. M. Davier, A. Hoecker, B. Malaescu, Z. Zhang, Reevaluation of the hadronic vacuum polarisation contributions to the Standard Model predictions of the muon $g - 2$ and $\alpha(m_Z^2)$ using newest hadronic cross-section data. *Eur. Phys. J. C* **77**, 827 (2017). <https://doi.org/10.1140/epjc/s10052-017-5161-6>. [arXiv:1706.09436](https://arxiv.org/abs/1706.09436)
31. M. Davier, A. Hoecker, B. Malaescu, Z. Zhang, A new evaluation of the hadronic vacuum polarisation contributions to the muon anomalous magnetic moment and to $\alpha(m_Z^2)$. *Eur. Phys. J. C* **80**, 241 (2020). <https://doi.org/10.1140/epjc/s10052-020-7792-2>. [arXiv:1908.00921](https://arxiv.org/abs/1908.00921). [Erratum: *Eur. Phys. J. C* **80**, 410 (2020)]
32. The Particle Data Group, Review of particle physics. *PTEP* **2022**, 083C01 (2022). <https://doi.org/10.1093/ptep/ptac097>
33. A. Keshavarzi, D. Nomura, T. Teubner, Muon $g - 2$ and $\alpha(M_Z^2)$: a new data-based analysis. *Phys. Rev. D* **97**, 114025 (2018). <https://doi.org/10.1103/PhysRevD.97.114025>. [arXiv:1802.02995](https://arxiv.org/abs/1802.02995)
34. A. Keshavarzi, D. Nomura, T. Teubner, $g - 2$ of charged leptons, $\alpha(M_Z^2)$, and the hyperfine splitting of muonium. *Phys. Rev. D* **101**, 014029 (2020). <https://doi.org/10.1103/PhysRevD.101.014029>. [arXiv:1911.00367](https://arxiv.org/abs/1911.00367)
35. C.K. Williams, C.E. Rasmussen, *Gaussian Processes for Machine Learning* (MIT press, Cambridge, 2006)

36. D.J. MacKay, *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, Cambridge, 2003)
37. D. Foreman-Mackey, E. Agol, S. Ambikasaran, R. Angus, Fast and scalable Gaussian process modeling with applications to astronomical time series. *Astron. J.* **154**, 220 (2017). <https://doi.org/10.3847/1538-3881/aa9332>. arXiv:1703.09710
38. S.O. Rice, Mathematical analysis of random noise. *Bell Syst Tech J* **23**, 282 (1944). <https://doi.org/10.1002/j.1538-7305.1944.tb00874.x>
39. R.B. Gramacy, H.K.H. Lee, Bayesian treed Gaussian process models with an application to computer modeling. *J. Am. Stat. Assoc.* **103**, 1119 (2008). <https://doi.org/10.1198/01621450800000689>. arXiv:0710.4536
40. R.B. Gramacy, *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences* (Chapman Hall/CRC, Boca Raton, 2020)
41. H.A. Chipman, E.I. George, R.E. McCulloch, Bayesian CART model search. *J. Am. Stat. Assoc.* **93**, 935 (1998). <https://doi.org/10.2307/2669832>
42. H.A. Chipman, E.I. George, R.E. McCulloch, Bayesian treed models. *Mach. Learn.* **48**, 299 (2002). <https://doi.org/10.1023/A:1013916107446>
43. D. Denison, N. Adams, C. Holmes, D. Hand, Bayesian partition modelling. *Comput. Stat. Data Anal.* **38**, 475 (2002). [https://doi.org/10.1016/S0167-9473\(01\)00073-1](https://doi.org/10.1016/S0167-9473(01)00073-1)
44. H.-M. Kim, B.K. Mallick, C.C. Holmes, Analyzing nonstationary spatial data using piecewise Gaussian processes. *J. Am. Stat. Assoc.* **100**, 653 (2005). <https://doi.org/10.1198/016214504000002014>
45. A. Lederer, A.J.O. Conejo, K. Maier, W. Xiao, J. Umlauf, S. Hirche, *Real-time regression with dividing local Gaussian processes*. arXiv:2006.09446
46. C.J. Paciorek, M.J. Schervish, Nonstationary covariance functions for gaussian process regression, in *Proceedings of the 16th International Conference on Neural Information Processing Systems, NIPS'03*, (Cambridge, MA, USA), pp. 273–280 (MIT Press, 2003)
47. A. Damianou, N.D. Lawrence, Deep Gaussian processes, in *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, ed. by C.M. Carvalho and P. Ravikumar, Proceedings of Machine Learning Research, vol. 31 (Scottsdale, Arizona, USA), pp. 207–215, PMLR (2013)
48. A.G. Wilson, Z. Hu, R. Salakhutdinov, E.P. Xing, Deep kernel learning, in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics* ed. by A. Gretton and C.C. Robert Proceedings of Machine Learning Research, vol. 51, (Cadiz, Spain), pp. 370–378, PMLR, 09–11 May, 2016. arXiv:1511.02222
49. V. Tolvanen, P. Jylänki, A. Vehtari, Expectation propagation for nonstationary heteroscedastic Gaussian process regression, in 2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6 (2014). <https://doi.org/10.1109/MLSP.2014.6958906>
50. M. Heinonen, H. Mannerström, J. Rousu, S. Kaski, H. Lähdesmäki, Non-stationary Gaussian process regression with Hamiltonian Monte Carlo, in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics* ed. by A. Gretton and C.C. Robert Proceedings of Machine Learning Research, vol. 51 (Cadiz, Spain), pp. 732–740, PMLR, 09–11 (2016)
51. A. Sauer, R.B. Gramacy, D. Higdon, Active learning for deep Gaussian process surrogates. *Technometrics* **1** (2022). <https://doi.org/10.1080/00401706.2021.2008505>. arXiv:2012.08015
52. A. Sauer, A. Cooper, R.B. Gramacy, Vecchia-approximated Deep Gaussian Processes for Computer Experiments. *J. Comput. Graph. Stat.* **1** (2022). <https://doi.org/10.1080/10618600.2022.2129662>. arXiv:2204.02904
53. P.J. Green, Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711 (1995). <https://doi.org/10.1093/biomet/82.4.711>
54. P.J. Green, D.I. Hastie, Reversible jump MCMC. *Genetics* **155**, 1391 (2009)
55. D.I. Hastie, P.J. Green, Model choice using reversible jump Markov chain Monte Carlo. *Stat. Neerl.* **66**, 309 (2012). <https://doi.org/10.1111/j.1467-9574.2012.00516.x>
56. S.A. Sisson, Transdimensional Markov chains: a decade of progress and future perspectives. *J. Am. Stat. Assoc.* **100**, 1077 (2005). <https://doi.org/10.1198/01621450500000664>
57. A. O'Hagan, Bayes-Hermite quadrature. *J. Stat. Plan. Inference* **29**, 245 (1991). [https://doi.org/10.1016/0378-3758\(91\)90002-V](https://doi.org/10.1016/0378-3758(91)90002-V)
58. Z. Ghahramani, C. Rasmussen, *Bayesian Monte Carlo*, in *Advances in Neural Information Processing Systems*, ed. by S. Becker, S. Thrun and K. Obermayer, vol. 15 (MIT Press, 2002)
59. M. Fisher, C. Oates, C. Powell, A. Teckentrup, *A Locally Adaptive Bayesian Cubature Method*, in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, ed. by S. Chiappa and R. Calandra, Proceedings of Machine Learning Research, vol. 108 pp. 1265–1275, PMLR, 26–28 Aug, 2020. arXiv:1910.02995
60. R. Jagadeeswaran, F.J. Hickernell, Fast automatic Bayesian cubature using lattice sampling. *Stat. Comput.* **29**, 1215 (2019). <https://doi.org/10.1007/s11222-019-09895-9>. arXiv:1809.09803
61. F.-X. Briol, C.J. Oates, M. Girolami, M.A. Osborne, D. Sejdinovic, Probabilistic integration: a role in statistical computation? *Stat. Sci.* **34**, 1 (2019). <https://doi.org/10.1214/18-ST660>. arXiv:1512.00933
62. H. Zhu, X. Liu, R. Kang, Z. Shen, S. Flaxman, F.-X. Briol, Bayesian Probabilistic Numerical Integration with Tree-Based Models, in *Advances in Neural Information Processing Systems*, ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan and H. Lin, vol. 33, pp. 5837–5849. Curran Associates, Inc. (2020)
63. D.J. MacKay, Information-based objective functions for active data selection. *Neural Comput.* **4**, 590 (1992). <https://doi.org/10.1162/neco.1992.4.4.590>
64. S. Seo, M. Wallat, T. Graepel and K. Obermayer, Gaussian process regression: Active data selection and test point rejection, in Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, vol. 3, pp. 241–246 (2000). <https://doi.org/10.1109/IJCNN.2000.861310>
65. P. Wei, X. Zhang, M. Beer, Adaptive experiment design for probabilistic integration. *Comput. Methods Appl. Mech. Eng.* **365**, 113035 (2020). <https://doi.org/10.1016/j.cma.2020.113035>
66. R.B. Gramacy, H.K.H. Lee, Adaptive design and analysis of super-computer experiments. *Technometrics* **51**, 130 (2009). <https://doi.org/10.1198/TECH.2009.0015>. arXiv:0805.4359
67. The Particle Data Group, Data files and plots of cross-sections and related quantities in the 2022 Review of Particle Physics (2023). <https://pdg.lbl.gov/2023/hadronic-xsections/hadron.html>
68. V.V. Ezhela, S.B. Lugovsky, O.V. Zenin, Hadronic part of the muon $g - 2$ estimated on the $\sigma_{\text{total}}^{2003}(e^+e^- \rightarrow \text{hadrons})$ evaluated data compilation. arXiv:hep-ph/0312114
69. R. Kumar, C. Carroll, A. Hartikainen, O. Martin, ArviZ a unified library for exploratory analysis of Bayesian models in Python. *J. Open Source Softw.* **4**, 1143 (2019). <https://doi.org/10.21105/joss.01143>
70. A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, P.-C. Bürkner, Rank-normalization, folding, and localization: an improved \hat{R} for assessing convergence of MCMC. *Bayesian Analysis* **16** (2019). <https://doi.org/10.1214/20-ba1221>. arXiv:1903.08008
71. Q. Li, Code and data associated with this paper (2023). <https://github.com/qiao688/TGP>

72. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1992)
73. M.W. Seeger, C.K.I. Williams, N.D. Lawrence, Fast Forward Selection to Speed Up Sparse Gaussian Process Regression, in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, ed. by C.M. Bishop and B.J. Frey, Proceedings of Machine Learning Research, vol. R4, pp. 254–261, PMLR, 03–06 (2003)
74. J. Quinero-Candela, C.E. Rasmussen, A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.* **6**, 1939 (2005)
75. E. Snelson, Z. Ghahramani, Local and global sparse Gaussian process approximations, in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, ed. by M. Meila and X. Shen, eds., Proceedings of Machine Learning Research, vol. 2 (San Juan, Puerto Rico), pp. 524–531, PMLR, 21–24 Mar (2007)
76. H. Liu, Y.-S. Ong, X. Shen, J. Cai, When Gaussian process meets big data: a review of scalable GPs. *IEEE Trans. Neural Netw. Learn. Syst.* **31**, 4405 (2020). <https://doi.org/10.1109/TNNLS.2019.2957109>
77. J. Hensman, N. Fusi, N.D. Lawrence, Gaussian processes for big data, in *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pp. 282–290 (2013). [arXiv:1309.6835](https://arxiv.org/abs/1309.6835)
78. J.-B. Delisle, N. Hara, D. Ségransan, Efficient modeling of correlated noise. *Astron. Astrophys.* **638**, A95 (2020). <https://doi.org/10.1051/0004-6361/201936906>. [arXiv:hep-ex/2004106](https://arxiv.org/abs/2004.106)