

GPU-Accelerated Modified Bessel Function of the Second Kind for Gaussian Processes

Zipei Geng¹, Sameh Abdulah², Ying Sun¹, Hatem Ltaief², David E. Keyes², Marc G. Genton¹

¹Statistics Program,

²Applied Mathematics and Computational Sciences (AMCS) Program,

^{1,2}King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.

^{1,2}{firstname.lastname}@kaust.edu.sa

Abstract—Modified Bessel functions of the second kind are widely used in physics, engineering, spatial statistics, and machine learning. Since contemporary scientific applications, including machine learning, rely on GPUs for acceleration, providing robust GPU-hosted implementations of special functions, such as the modified Bessel function, is crucial for performance. Existing implementations of the modified Bessel function of the second kind rely on CPUs and have limited coverage of the full range of values needed in some applications. In this work, we present a robust implementation of the modified Bessel function of the second kind on GPUs, eliminating the dependence on the CPU host. We cover a range of values commonly used in real applications, providing high accuracy compared to common libraries like the GNU Scientific Library (GSL) when referenced to Mathematica as the authority. Our GPU-accelerated approach also demonstrates a 2.68X performance improvement using a single A100 GPU compared to the GSL on 40-core Intel Cascade Lake CPUs. Our implementation is integrated into *ExaGeoStat*, the HPC framework for Gaussian process modeling, where the modified Bessel function of the second kind is required by the Matérn covariance function in generating covariance matrices. We accelerate the matrix generation process in *ExaGeoStat* by up to 12.62X with four A100 GPUs while maintaining almost the same accuracy for modeling and prediction operations using synthetic and real datasets.

Index Terms—Modified Bessel function of the second kind, GPU implementation, Tile-based matrix computations, Gaussian processes.

I. INTRODUCTION

The modified Bessel function of the second kind, denoted as $K_\nu(x)$, where ν represents the order and x the argument, arises in a wide range of applications across mathematics, physics, engineering, spatial statistics, and machine learning. For instance, in physics and engineering, $K_\nu(x)$ frequently appears in solving PDEs in cylindrical symmetry with radially decaying behavior. Examples include heat conduction, damped wave propagation [1], electromagnetic fields in cylindrical waveguides [2], and Schrödinger’s equation in axisymmetry [3]. Similarly, in Gaussian processes, $K_\nu(x)$ is used in the Matérn kernel function, which is widely used to construct correlation matrices that capture spatial relationships between locations [4]. These diverse

applications emphasize the importance of efficient computation of the $K_\nu(x)$ function.

In the literature, three methods are widely employed to compute the modified Bessel function of the second kind across a broad range of values for x and ν [5]: series expansions [6], continued fractions [7], and asymptotic expansions [8]. The first expresses $K_\nu(x)$ as an infinite series that involves powers of x . This approach is particularly effective for small x , where truncating the series to a finite number of terms provides a highly accurate approximation. The continued fractions method represents the Bessel function as an infinite recursive fraction. This method is suitable for larger x than is the series expansion, but it remains computationally intensive. An asymptotic expansion approximates the Bessel function by capturing its dominant behavior as x becomes very large. By neglecting lower-order terms, this method efficiently describes the primary growth trend. However, it lacks accuracy for small x , where such approximations fail to converge effectively. Combining these three methods and other less common approaches can effectively cover a wider range of x and ν values. Nevertheless, these techniques are computationally demanding and inherently sequential, posing challenges for parallelization on modern architectures, such as GPUs.

In this paper, we adopt an integration-based method proposed by [5] to compute $K_\nu(x)$ for real order $\nu \in \mathbb{R}$, denoted by BESSELK, and optimize it for GPU implementation. This approach, implemented in CUDA, is well suited for parallel execution, as it represents the BESSELK function using quadrature over finite intervals. The key advantages of this method, beyond its parallelizability, are its flexibility and its ability to efficiently cover a reasonable range of x and ν values. Our implementation has been integrated into *ExaGeoStat* [9], a scalable geospatial data modeling and prediction framework optimized for manycore systems, including modern GPUs. *ExaGeoStat* possesses three functionalities: synthetic data generation, modeling, and prediction. Each of these operations requires, at a minimum, the generation of a covariance matrix

of size $N \times N$, where N represents the number of spatial locations in its simplest form. The Matérn kernel (or covariance function), which involves the BESSELK function, is often used in Gaussian processes to construct this covariance matrix.

Our contributions are summarized as follows:

- We present a highly efficient GPU implementation of the BESSELK function, which addresses a critical performance bottleneck encountered in numerous scientific applications.
- We accelerate an integration-based algorithm using a unified integral bound for BESSELK, enabling efficient handling of a practical range of x and ν commonly encountered in Gaussian processes on the GPU.
- We combine the accelerated integration-based algorithm with a series expansion to accurately compute the BESSELK function for $x < 0.1$, where the integral-based approach specifically underperforms, making our solution more comprehensive.
- Our refined algorithm combines series expansion and integration, both of which are established methods to approximate BesselK. However, to the best of our knowledge, we are the first to deploy these approaches on massively parallel GPU hardware accelerators.
- We integrate the refined algorithm into the covariance matrix generation step of *ExaGeoStat* and evaluate its performance and accuracy against existing CPU-based implementations, such as the GNU Scientific Library.
- We assess the accuracy of the proposed algorithm to compute the BESSELK function against existing approaches. Additionally, we evaluate its accuracy and overall impact within the *ExaGeoStat* software pipeline using synthetic datasets and a real-world wind speed dataset for climate and weather modeling applications.

II. RELATED WORK

The modified Bessel function of the second kind, $K_\nu(x)$ or BESSELK, arises in the separation of variables solutions of partial differential equations [10]–[12]. A key feature of BESSELK is its *rapid decay at infinity*. The function diminishes exponentially as $x \rightarrow \infty$, providing a basis for solutions that vanish at large distances. Applications include heat conduction [13], wave propagation [14], and stochastic processes [15].

BESSELK also arises in the widely used Matérn kernel, a covariance function for Gaussian processes. The Matérn kernel finds extensive applications across various fields. In spatial statistics, it is used to model spatial correlations [4]. In machine learning, it is particularly valuable for Gaussian process regression and Bayesian

optimization [16]. In signal processing, it helps capture correlations in time series and spectral analysis [17].

At least six different methods for BESSELK can be identified in the literature; however, not all of these methods are adopted in existing tools and libraries. (1) *Series Expansions* [6]: $K_\nu(x)$ can be expressed as an infinite series in powers of x , which is truncated to a finite number of terms when x is small. Most existing libraries utilize series expansions for small x , including MATLAB, Mathematica, GSL, SciPy, and Maple. (2) *Continued Fractions* [7]: $K_\nu(x)$ can be represented as an infinite fraction with a recursive structure, particularly suitable when x is moderate or large. This method is also used in most existing libraries for moderate x values. (3) *Asymptotic Expansions* [8]: This method focuses on the dominant behavior of x as it becomes large, ignoring lower-order terms, which reduces the accuracy for small x . This method is also used in most existing libraries for large x values. (4) *Integral Representations* [5], [18]: $K_\nu(x)$ is expressed as integrals. This method is flexible and can work with small or large values of x . Some libraries that rely on this method include Mathematica, SciPy, and GSL. (5) *Polynomial Fitting* [19]: Polynomial fitting approximates $K_\nu(x)$, by fitting polynomials to precomputed values of the function over restricted ranges of x and ν . (6) *Differential Equation Solvers* [20]: Numerical methods for solving differential equations, such as finite difference schemes and Runge-Kutta methods [21], can be employed to compute $K_\nu(x)$. These solvers approximate the solution of $K_\nu(x)$ by discretizing the domain and applying iterative algorithms to solve the governing differential equation.

III. BACKGROUND

This section addresses the challenges in the modeling of Gaussian processes and highlights the importance of accelerating the computations of the BESSELK functions to enhance the modeling procedure. Furthermore, we provide an overview of the *ExaGeoStat* software and its implementation, as a testbed to evaluate our novel BESSELK implementation.

A. Gaussian Processes and Matérn Kernel

Gaussian process (GP) modeling is a probabilistic machine learning approach that defines distributions over functions and distinguished by its mean function $m(\mathbf{x})$, and covariance function $C(\mathbf{x}, \mathbf{x}')$, where \mathbf{x}, \mathbf{x}' are locations. A commonly used method in Gaussian process modeling is the maximum likelihood estimation (MLE). In MLE, an optimization process iterates over a given log-likelihood function to estimate statistical parameters that describe the underlying data. This is achieved by constructing a positive definite covariance matrix that

captures the correlation between different observations. The log-likelihood function $\mathcal{L}(\boldsymbol{\theta})$ is represented as:

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2} [N \log(2\pi) + \log(|\boldsymbol{\Sigma}(\boldsymbol{\theta})|) + \mathbf{z}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{z}],$$

where $\boldsymbol{\theta}$ is a set of parameters to be fit, N is the number of observations, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the parameterized covariance matrix, $|\boldsymbol{\Sigma}(\boldsymbol{\theta})|$ is its determinant, and \mathbf{z} is the vector of observed data. The Matérn function serves as a fundamental component in spatial statistics, representing a significant domain where Gaussian processes demonstrate their utility, owing to its adaptability in characterizing spatial correlation structures. Its parametric flexibility enables precise modeling of spatial relationships across diverse applications and datasets [4].

The Matérn function can be represented as:

$$C(\mathbf{x}, \mathbf{x}') = \mathcal{M}(r; \boldsymbol{\theta}) = \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)} \left(\frac{r}{\beta}\right)^\nu K_\nu\left(\frac{r}{\beta}\right),$$

where $r = \|\mathbf{x} - \mathbf{x}'\|$ is the distance between two spatial locations and $\boldsymbol{\theta} = (\sigma^2, \beta, \nu)^\top$, $\nu > 0$ is a parameter that controls the smoothness of the function, β is the length scale parameter, σ^2 is the variance, $\Gamma(\nu)$ is the Gamma function, and $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν (BESSELK(x, ν)).

The BESSELK is the core in the generation of the covariance matrix when using the Matérn function, which is required not only during the modeling process but also for prediction tasks and the generation of synthetic datasets [22], [23].

B. ExaGeoStat: A Parallel Tile-Based Framework for Geospatial Data Analysis

ExaGeoStat is a high-performance software package for large-scale climate and environmental geostatistics [9]. It evaluates the log-likelihood function for spatial datasets using a range of covariance models, including the Matérn covariance function. This enables efficient parameter estimation and prediction for large-scale spatial datasets by leveraging state-of-the-art dense linear algebra libraries (e.g., CHAMELEON [24], and DPLASMA [25]) and runtime systems (e.g., StarPU [26], and PaRSEC [27]). *ExaGeoStat* achieves high performance across diverse hardware architectures, including multicore CPUs, GPUs, and distributed systems. The software supports both exact and approximate computations [28]–[31].

ExaGeoStat relies on tile-based algorithms to effectively leverage the underlying runtime system to distribute tasks across available hardware resources. The runtime system efficiently schedules the processing of matrix tiles across computational units, maximizing performance and resource utilization. All executions are conducted exclusively on GPUs to further optimize the performance of *ExaGeoStat*.

IV. ALGORITHM FOR BESSELK

Our BESSELK algorithm is built to compute a reasonable range of x , and ν that covers a wide spectrum of applications builds upon two existing methods: Temme's series expansion and Takekawa's algorithm, reviewed, respectively, in the following two subsections.

A. Temme's series expansion

Most existing libraries rely on series expansions to compute the BESSELK function when x is small [6]. Series expansions offer an efficient and accurate approximation in this regime, with the series expressed as:

$$K_\nu(x) = \sum_{k=0}^{\infty} c_k f_k, \quad K_{\nu+1}(x) = \frac{2}{x} \sum_{k=0}^{\infty} c_k h_k, \quad (1)$$

where

$$c_k = \frac{(x^2/4)^k}{k!}, \quad h_k = -k f_k + p_k,$$

and p_k and q_k are recurrence relations defined as:

$$p_k = \frac{p_{k-1}}{k - \nu}, \quad q_k = \frac{q_{k-1}}{k + \nu}.$$

The term f_k is computed iteratively:

$$f_k = \frac{k f_{k-1} + p_{k-1} + q_{k-1}}{k^2 - \nu^2}.$$

The recurrence relations are initialized with:

$$p_0 = \frac{1}{2} \left(\frac{x}{2}\right)^{-\nu} \Gamma(1+\nu), \quad q_0 = \frac{1}{2} \left(\frac{x}{2}\right)^\nu \Gamma(1-\nu), \quad (2)$$

$$f_0 = \frac{\nu\pi}{\sin(\nu\pi)} \left[\cosh(\sigma) \Gamma_1(\nu) + \frac{\sinh(\sigma)}{\sigma} \ln\left(\frac{2}{x}\right) \Gamma_2(\nu) \right], \quad (3)$$

where σ and $\Gamma_1(\nu)$, $\Gamma_2(\nu)$ are precomputed constants for the expansion.

The original formulation of Temme's series expansion in [6] does not provide sufficiently stable or accurate results for $\nu \geq 1.5$. The direct evaluation of p_0, q_0 , and f_0 for large orders of ν increases the error of this approximation method. To address these limitations and improve numerical stability and accuracy, Campbell [32] proposed leveraging the recurrence relation for large-order $\nu = \mu + M$ (where $M = \lfloor \nu + 0.5 \rfloor$) of $K_\nu(x)$.

This approach uses starting values $K_\mu(x)$ and $K_{\mu+1}(x)$, with $-\frac{1}{2} \leq \mu < \frac{1}{2}$, and applies the forward recurrence relation:

$$K_{\eta+1}(x) = \left(\frac{2\eta}{x}\right) K_\eta(x) + K_{\eta-1}(x). \quad (4)$$

We employ Temme's series expansion and the recurrence relation to evaluate the BESSELK function in the small- x regime ($x < 0.1$).

B. Integral-based Algorithm (Takekawa's Approach)

In [5], Takekawa introduced a method to compute the BESSELK function using an integral approach. $K_\nu(x)$ is represented by the following integral from [33]:

$$K_\nu(x) = \int_0^\infty e^{-x \cosh(t)} \cosh(\nu t) dt \triangleq \int_0^\infty f_{\nu,x}(t) dt, \quad (5)$$

where $x > 0$ and $\nu \in \mathbb{R}$. This representation is computationally intensive but provides accurate results, especially for small and moderate x .

Takekawa worked with the logarithm of the integrand $e^{-x \cosh(t)} \cosh(\nu t)$. Namely,

$$g_{\nu,x}(t) = \log \cosh(\nu t) - x \cosh(t). \quad (6)$$

The first-order, and second-order derivatives of $g_{\nu,x}(t)$ with respect to t are given by:

$$g'_{\nu,x}(t) = \nu \tanh(\nu t) - x \sinh(t),$$

$$g''_{\nu,x}(t) = \nu^2 \operatorname{sech}^2(\nu t) - x \cosh(t).$$

At $t = 0$, we have $g_{\nu,x}(0) = -x$ and $g'_{\nu,x}(0) = 0$. Consequently, if $\nu^2 \leq x$, the function $g_{\nu,x}(t)$ will always decrease, implying that the maximum value of the function in Equation (6) occurs at $t = 0$, as $g''_{\nu,x}(0) \leq 0$. In contrast, if $\nu^2 > x$, the function in Equation (6) will reach its maximum at some $t \geq 0$, since $g''_{\nu,x}(0) > 0$. More details are provided in [5].

To integrate Equation (5), the region where the maximum value of t , denoted as t_{\max} , is located, can be defined as the region where $f_{\nu,x}(t) \geq \epsilon_{\text{machine}} f_{\nu,x}(t_{\max})$, where $\epsilon_{\text{machine}}$ is the machine epsilon. This region can be expressed as interval $[t_0, t_1]$, defined as:

$$[t_0, t_1] = \{t \mid g_{\nu,x}(t) \geq \log(\epsilon_{\text{machine}}) + g_{\nu,x}(t_{\max})\}. \quad (7)$$

Takekawa defines the integral range by determining t_{\max} from Equation (6). If $\nu^2 \leq x$, then $t_{\max} = 0$. Otherwise, t_{\max} can be found by searching within a specific range of the function $g_{\nu,x}(t)$. Since $g'_{\nu,x}(t_{\max}) = 0$ and $g'_{\nu,x}(t) < 0$ for $t > t_{\max}$, the range can be defined as $[2^{m-1}, 2^m]$, where m is the smallest value such that $g'_{\nu,x}(2^m) < 0$. For a detailed explanation, refer to the *FINDRANGE* algorithm in [5]. Afterward, the t_{\max} value can be obtained using binary search and Newton methods; refer to the *FINDZERO* algorithm in [5].

The integration range is also determined using the *FINDZERO* algorithm. To compute t_0 (the lower bound of integration), if $\nu^2 \leq x$, then $t_0 = 0$. Otherwise, the *FINDZERO* algorithm is applied to find t_0 within the range $[t_0, t_{\max}]$. For t_1 , the *FINDZERO* algorithm is used similarly to search within the range $[t_{\max} + 2^{m-1}, t_{\max} + 2^m]$.

After determining t_0 and t_1 , integration is performed over the range using a fixed number b of bins. To

ensure numerical stability, the `log_sum_exp` function is applied to $g_{\nu,x}(t_m)$, resulting in:

$$\log K_\nu(x) \approx g_{\nu,x}(t_{\max}) + \log \sum_{m=0}^b h \exp \left\{ c_m (g_{\nu,x}(t_m) - g_{\nu,x}(t_{\max})) \right\}. \quad (8)$$

Here, the parameters are defined as:

$$h = \frac{t_1 - t_0}{b}, \quad t_m = t_0 + mh, \\ c_0 = c_n = \frac{1}{2}, \quad c_m = 1 \quad (m = 1, \dots, b-1).$$

$K_\nu(x)$ can be obtained by taking the exponential.

C. The Proposed Refined Algorithm

Existing numerical libraries, such as MATLAB, SciPy, and Boost C++, do not offer GPU support to evaluate the BESSELK function, resulting in a significant performance bottleneck for GPU-accelerated scientific applications that depend on this function. Performing BESSELK computations on the CPU while executing other tasks on the GPU is highly time-consuming due to the overhead of data transfers between the CPU and GPU and the GPU's superior parallel processing capabilities compared to the CPU. Although Plesner et al. [34] recently introduced a GPU-accelerated library for BESSELK evaluation, its performance gains are limited. Their implementation outperforms GSL only within the parameter range $(x, \nu) \in (150, 4000] \times (150, 4000]$ and exhibits a relatively limited performance outside of this range compared to GSL.

To address this gap, we propose a novel approach that advances the state-of-the-art in GPU-accelerated BESSELK computation, delivering improved performance across a reasonable parameter space. We adopt the quadrature-based algorithm proposed by Takekawa [5]. Although Takekawa's algorithm demonstrates substantial accuracy for large values of x and ν , it notably lacks discussion of cases where $x < 0.1$, a range frequently encountered in applications such as spatial statistics. Our analysis reveals that within this range, the integration algorithm exhibits a significant loss of accuracy. Figure 1 shows a heatmap of the relative error in Takekawa's approach for $x < 0.1$, compared to referential results obtained using Mathematica [35]. Following Takekawa in [5], we compute the relative error (RE) at each x and ν as:

$$\text{RE} = \log_{10} \left(1 + \frac{|\text{Mathematica's output} - \text{output}|}{\epsilon_{\text{machine}}} \right)$$

where $\epsilon_{\text{machine}} = 2^{-52} \approx 2.22 \times 10^{-16}$ for double-precision numbers.

We propose a novel algorithm for the BESSELK function that extends the range of x values beyond those

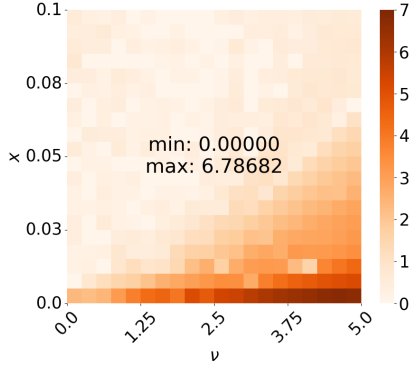


Fig. 1: Relative error of Takekawa's algorithm vs Mathematica for $(\nu, x) \in [0.001, 5] \times [0.001, 0.1]$.

supported by Takekawa's method. We refer to this enhanced approach as the *Refined Algorithm*. Furthermore, we provide an efficient GPU-based implementation for generating large matrices using BESSELK within the *ExaGeoStat* software that relies on the *StarPU* runtime system. Our key contributions to improving Takekawa's algorithm are summarized as follows:

- 1) Our algorithm simplifies the integral in Equation (5) by setting the lower bound to 0, as reducing the range by computing t_0 and t_1 is more computationally expensive than extending the range on the GPU.
- 2) The upper bound is set to a maximum value determined through an empirical bound finder, ensuring applicability across all x and ν ranges.
- 3) Instead of using the *FINDZERO* algorithm to find a global t_{\max} , as in Takekawa's method, we compute a local t_{\max} for each division within the integral range, enabling faster computation.
- 4) To improve accuracy, we increase the number of bins b in the integral, which offsets the expanded range $[t_0, t_1]$ and leverages GPU computational power with minimal performance impact.
- 5) For $x < 0.1$, we combine the integral algorithm with Temme's expansion, reducing the relative error from 6.78682 in Takekawa's method to 0.73180 in the refined algorithm, provided a sufficient number of bins.

Algorithm 1 Empirical Upper Bound Finding

- 1: **Given:** $\mathcal{X} \times \mathcal{V} = [0, 140] \times (0, 20]$, $\text{MBK}(x, \nu)$ the reference level of LOGBESSELK using Mathematica, and $\text{RBK}(x, \nu) := \log \left[\int_0^L e^{-x \cosh(t)} \cosh(\nu t) dt \right]$, where the integration follows the Equation (8).
 - 2: **for all** $(x, \nu) \in \mathcal{X} \times \mathcal{V}$ **do**
 - 3: $\text{AE}(x, \nu) \leftarrow |\text{MBK}(x, \nu) - \text{RBK}(x, \nu)|$ (Absolute Error for (x, ν))
 - 4: $t_1 \leftarrow \min L$ s.t. $\max_{x, \nu} \text{AE}(x, \nu) \leq 10^{-9}$
-

For $x \in [0.1, \infty)$, we use a refined version of the integration algorithm proposed in [5] to evaluate BESSELK. Instead of determining the integration range

$[t_0, t_1]$ dynamically based on x and ν , we fix $t_0 = 0$ and $t_1 = 9$. We set $t_0 = 0$ to match the starting point of the integral, and through the empirical upper bound finding algorithm, we establish $t_1 = 9$ as the optimal endpoint for parameters within $(x, \nu) \in [0, 140] \times (0, 20]$, which corresponds to the expected parameter range in geospatial applications involving the Matérn kernel. To elaborate on, in geospatial and machine learning studies, we can always rescale the 2D plane inside a unit square, whose maximum distance is $\sqrt{2}$. The typical starting point of optimizing the β parameter in the Matérn kernel is 0.01, which gives us the maximum x we have $\sqrt{2}/0.01 \approx 140$. We are choosing $x \in [0, 140]$. The smoothness parameter ν controls the smoothness of the data and has mathematical interpretation as the m -th differentiability for integers $m < \nu$. The case $\nu > 20$ can be approximated by the squared exponential kernel. Thus, choosing this study region is reasonable. Empirically finding the upper bound is shown in Algorithm 1. This broader integration range may slightly increase computational cost but improves GPU efficiency by avoiding performance-degrading conditional branching. Besides, this unified upper bound, no matter what (x, ν) , largely decreased the computational burden compared to the original method, which determines the lower and upper bound for each (x, ν) pair using Newton-based zero-finders.

Algorithm 2 provides a detailed explanation of the steps of our proposed algorithm. The inputs are the pair (x, ν) to compute $K_\nu(x)$. To define b , the number of bins, increasing it can improve accuracy, but at the cost of reduced performance, as the algorithm must identify the local maximum point t_{\max} for each division. However, we observed that fixing the number of bins to 40 provides a balance, achieving an accuracy threshold that ensures computational stability across different values of x and ν . In Algorithm 2, Temme's expansion method is used to compute the Bessel function if $x \leq 0.1$ (lines 3-7); otherwise, use Equation (8) to compute the Bessel function using b bins (lines 8-13).

Algorithm 3 presents the pseudocode for the CUDA algorithm used to generate a single tile of the covariance matrix based on the Matérn covariance function and the BESSELK function, using the refined algorithm. Using the *StarPU* runtime system, the matrix is partitioned into smaller tiles and each tile is assigned to a different GPU to dynamically compute the full covariance matrix, while handling the data transfer from host-to-device and device-to-host. Memory allocation is handled via *StarPU* to optimize data transfer performance between the CPU (host) and the GPU. In lines 5-13, each GPU thread processes a single tile value to compute the BESSELK function based on the corresponding x value, as described in Algorithm 2.

Algorithm 2 Refined Algorithm (BESSELK(x, ν))

```

1: Input:  $(x, \nu) \in \mathcal{X} \times \mathcal{V}$ , where  $\mathcal{X} \times \mathcal{V}$  is the problem
   region of evaluation of BESSELK
2: Given:  $b$  number of bins for numerical integration
3: if  $0 \leq x < 0.1$  and  $\nu \in \mathcal{V}$  then
4:   Set  $M = \lfloor \nu + 0.5 \rfloor$  and use  $\mu = \nu - M$ 
5:   Use Equation (2) and Equation (3) to initialize
    $p_0, q_0, f_0$  for smoothness  $\mu$ 
6:   Set 15000 instead of  $\infty$  for the sum of Temme's series
   for  $K_\mu(x)$  in Equation (1)
7:   Use the recurrence relation
      
$$K_{\eta+1}(x) = \left(\frac{2\eta}{x}\right) K_\eta(x) + K_{\eta-1}(x)$$

   until  $\eta + 1 = \nu$  to acquire  $K_\nu(x)$ 
8: else  $x \in \mathcal{X} \setminus [0, 0.1)$  and  $\nu \in \mathcal{V}$ 
9:    $t_0 \leftarrow \text{LB}$  (fixed to 0)
10:   $t_1 \leftarrow \text{UB}$  (empirical upper bound for all  $(x, \nu)$ )
11:   $t_{\text{lmax}} \leftarrow \max_{i=0, \dots, b} t_i$ 
12:  Use
      
$$\log K_\nu(x) \leftarrow g_{\nu, x}(t_{\text{lmax}}) +$$

      
$$\log \sum_{m=0}^b h \exp \left\{ c_m (g_{\nu, x}(t_m) - g_{\nu, x}(t_{\text{lmax}})) \right\},$$

      where  $h, t_m, c_m$  are given in Equation (8) and  $g_{\nu, x}(\cdot)$ 
      given in Equation (6)
13:   $K_\nu(x) \leftarrow \exp(\log(K_\nu(x)))$ 
14: Output: BESSELK( $x, \nu$ )  $\leftarrow K_\nu(x)$ 

```

Algorithm 3 GPU Single-Tile Matérn Covariance Generation Algorithm

```

1: function GENERATEMATÉRNCOVARIANCE( $\ell_x^1, \ell_y^1, \ell_x^2, \ell_y^2,$ 
    $\sigma, \beta, \nu, m, n$ ) where  $\ell^1$  and  $\ell^2$  represents location vectors.
2:   Initialize CUDA grid with dimension  $(m+8-1//m) * (n+8-1//n)$ 
   and block dimensions  $8 \times 8$  (which means 64 threads for each block)
3:   Use starpu_malloc() for efficient CUDA memory allocation,
   enabling fast CPU-GPU and GPU-GPU transfers.
4:   Copy location vectors and parameters to GPU memory
5:   for each thread  $(i, j)$  in parallel do
6:     if  $i < m$  and  $j < n$  then
7:        $c \leftarrow \sigma^2 / (2^{\nu-1} \Gamma(\nu))$ 
8:        $d \leftarrow \sqrt{(\ell_x^2[j] - \ell_x^1[i])^2 + (\ell_y^2[j] - \ell_y^1[i])^2}$ 
9:        $r \leftarrow d/\beta$ 
10:      if  $r = 0$  then
11:         $A[i + j \times m] \leftarrow \sigma^2$ 
12:      else
13:         $A[i + j \times m] \leftarrow c \cdot r^\nu \cdot \text{BESSELK}(r, \nu)$ 
14:      Transfer back the generated submatrices and parameters to CPU
15:    Synchronize CUDA stream
16:  Free the CUDA memory and destroy the CUDA stream

```

D. GPU Optimizations on Algorithm 2

Algorithm 2 is written for CUDA where threads are collected in blocks of threads; 64 threads in our case gained good performance. The threads in a block are collected in warps of 32 threads, where each warp is used for the Single Instruction Multiple Threads (SIMT) execution model [34]. When naively parallelized on GPUs, the i -th thread computes $K_\nu(x)$ for the i -th input value (x_i, ν_i) .

Specifically, the covariance matrix we target involves computing the modified Bessel function for r/β , where r represents the distance between two spatial locations, and β is a fixed parameter during a single generation operation. In our implementation, we ordered the spatial locations using Morton's ordering [36] to ensure that distances within the same block remain close and consistently follow a single branch of the if-else condition in the CUDA implementation since we are operating in a tile-based manner.

V. EXPERIMENTAL RESULTS

In this section, we evaluate the implementation of the refined BESSELK algorithm on the GPU within the *ExaGeoStat* software, focusing on its impact in accelerating the computation of individual matrix elements when millions to trillions of elements are computed. The experiments focus on four main objectives: (1) Assessing the accuracy of computing BESSELK for specific x and ν values; (2) Analyzing the overall accuracy of spatial statistical modeling using a synthetic dataset, emphasizing our GPU-based implementation for matrix generation across different spatial correlation levels; (3) Validating the accuracy of the proposed implementation in modeling real datasets within the context of climate and weather applications; (4) Evaluating the performance of full covariance matrix generation within *ExaGeoStat*, using both single and multiple GPUs.

The implementation associated with this research, which functions independently from the *ExaGeoStat* framework and focuses exclusively on optimizing BESSELK for the CUDA architecture, is available for public access at: <https://github.com/stsds/CuBesselK>

A. Experimental Configuration

The results for the overall accuracy of the spatial statistical modeling (Figures 4 – 7) are generated using NVIDIA GV100 32GB with Intel Cascade Lake 256GB CPU, each with 100 repetitions. The real dataset application follows the same settings. Performance assessments (Figures 9 – 12) use five repetitions for benchmarking. We use NVIDIA GV100 32GB and NVIDIA A100 80GB GPUs with Intel Cascade Lake CPU. For *ExaGeoStat*, we use CUDA 11.8, gcc 11.2.1, CMake 3.24.2, OpenMPI 4.1.4, OneAPI 2022.3, and StarPU 1.3.11.

B. Relative Error Analysis Against Mathematica

For accuracy assessment, existing work on implementing the BESSELK function often uses Mathematica's `BesselK[nu, x]` as the benchmark. In this study, we compare the performance of the GSL library, Takekawa's algorithm, and the refined algorithm against Mathematica. The relative errors for various values of x and ν are presented as a heatmap. The heatmap spans the region $(\nu, x) \in [0.001, 20] \times [0.001, 140]$, which adequately covers the parameter range relevant to Gaussian processes.

Figure 2 presents three heatmaps for the target region using the three implementations. As shown, the relative errors of the LOGBESSELK in Takekawa's algorithm are larger than those of GSL and the refined algorithm, with a maximum error of 6.54807. This is mainly due to the uncovered region where $x < 0.1$. GSL and the refined algorithm exhibit very close relative errors, with values of 1.67896 and 1.65466, respectively. To highlight the advantages of our method over Takekawa's method for $x \leq 0.1$, we further zoom into the region $(\nu, x) \in [0.001, 5] \times [0.001, 0.1]$ in Figure 3.

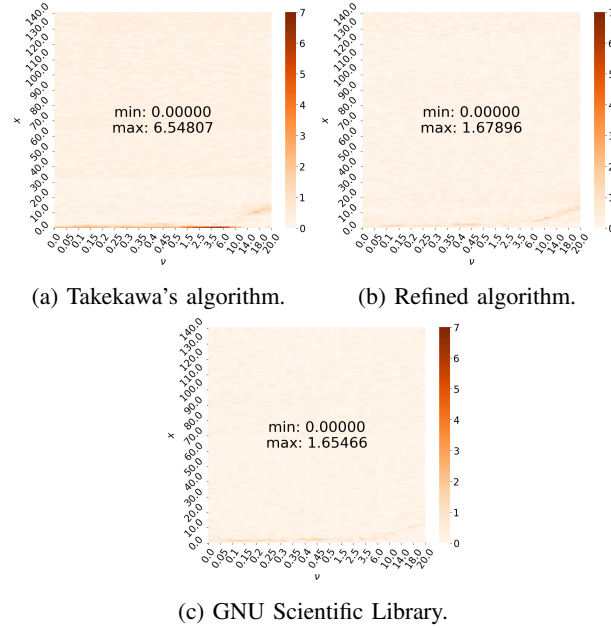


Fig. 2: LOGBESSELK accuracy comparisons using heatmap for $(\nu, x) \in [0.001, 20] \times [0.001, 140]$.

C. Spatial Data Modeling Accuracy within ExaGeoStat

Estimating the relative error of the Bessel function for each pair (x, ν) is crucial to evaluate the effectiveness of a given implementation. However, beyond this, the function BESSELK, which is buried within a kernel to generate a matrix, can affect the accuracy of subsequent matrix operations if the accumulated error from different calculations becomes significant. We integrate our

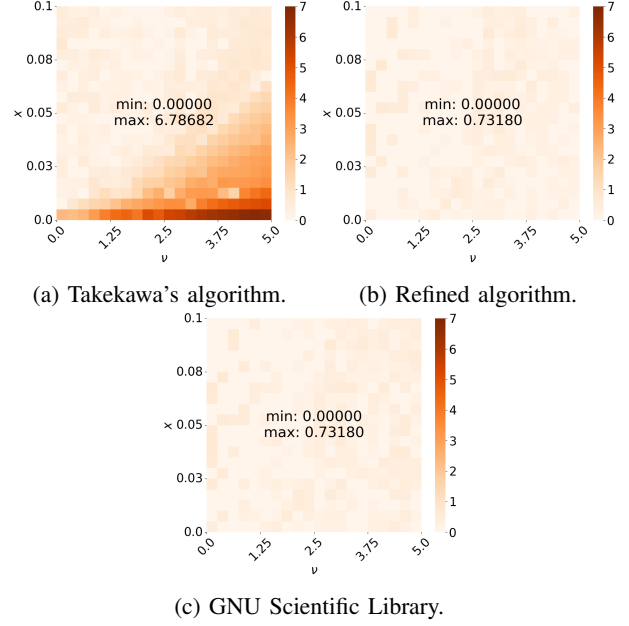


Fig. 3: LOGBESSELK accuracy comparisons using heatmap for $(\nu, x) \in [0.001, 5] \times [0.001, 0.1]$.

implementation into *ExaGeoStat* to generate the Matérn kernel covariance matrix to assess this point. We use the modeling process in *ExaGeoStat*, specifically MLE with gradient-free optimization, which requires multiple iterations to converge and generate a set of estimates of the parameters σ^2 , β and ν that effectively describe the underlying field.

The accuracy of the MLE operation is typically evaluated using simulations in which synthetic data are generated [37]. We use synthetic data generated within a 2D space, as described in [38]. We assess the accuracy of the modeling results across three representative scenarios commonly encountered in spatial statistics. These scenarios were characterized by different levels of spatial correlation: weak ($\beta = 0.03$ and $\beta = 0.025$), medium ($\beta = 0.1$ and $\beta = 0.075$) and strong ($\beta = 0.3$ and $\beta = 0.2$), for two different levels of smoothness, rough field ($\nu = 0.5$) and smooth field ($\nu = 1$), respectively. The variance was fixed at $\sigma^2 = 1$ in all experiments.

Figures 4a – 4f provide a comparison of parameter estimation using MLE, taking advantage of both GSL and the proposed refined algorithm. The analysis, conducted across three correlation levels (weak, medium, and strong), includes boxplots that illustrate the estimation of three key parameters (σ^2 , β , ν) and iteration counts.

In the weak correlation scenario (Figure 4a), both implementations achieve comparable accuracy in parameter estimation. Furthermore, the average iteration count for the refined algorithm is similar to that of GSL, indicating

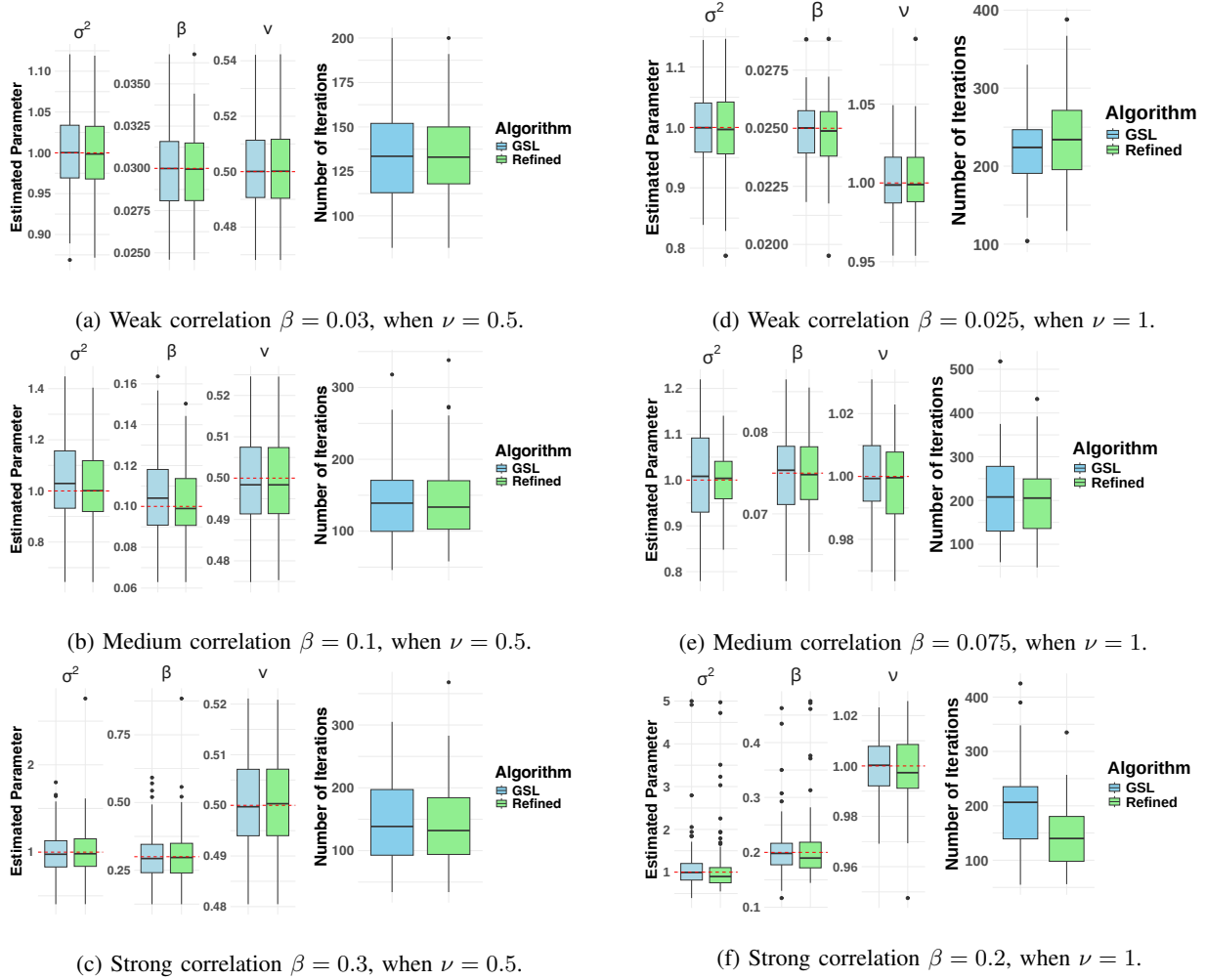


Fig. 4: Boxplots of MLE optimization results over 100 replicas comparing GSL (CPU) and refined algorithm (GPU). Left column: $\nu = 0.5$ cases. Right column: $\nu = 1$ cases. All plots show parameter estimates (σ^2 , β , ν) and iteration counts with red dashed lines indicating true values of parameters.

a comparable computational efficiency. The medium correlation scenario (Figure 4b) demonstrates increased variability in parameter estimation. Both implementations maintain comparable accuracy, but the refined algorithm exhibits slightly more consistent estimates across all parameters. For strong spatial correlation (Figure 4c), the most challenging scenario, both implementations exhibit wider parameter distributions, reflecting the increased difficulty of the estimation process. The refined algorithm achieves comparable accuracy and number of iterations.

In the smoother case where $\nu = 1.0$, Figures 4d, 4e, and 4f demonstrate that the refined algorithm achieves an accuracy similar to GSL. However, it shows a slightly higher bias between the median and ground truth in the strong correlation scenario, which remains within an acceptable range. This is expected, as a larger ν

makes the BESSELK approximation more sensitive to the number of bins used for numerical integration. However, the refined algorithm requires significantly fewer iterations than GSL. We also experimented with a strong correlation scenario using $b = 40$ instead of 16. The results are shown in Figure 5. The refined algorithm achieves a level of accuracy similar to that of GSL, and the bias in parameter estimation is eliminated. These results indicate that the refined algorithm largely preserves accuracy compared to GSL while achieving improved computational efficiency. Despite requiring a similar number of optimization iterations, each iteration is significantly faster, as demonstrated in the next section. Furthermore, consistent performance across varying correlation strengths highlights the robustness of the refined algorithm for spatial statistics and Gaussian process applications. By comparing Figure 4f and Figure 5, it

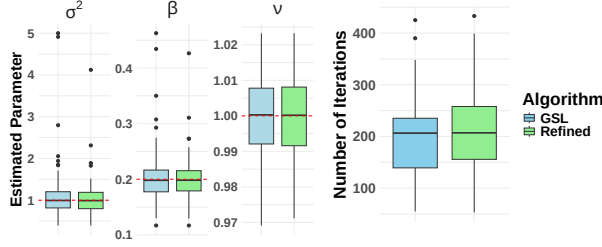


Fig. 5: Boxplots of MLE optimization results over 100 replicas using GSL on CPU and the refined algorithm on GPU when $\nu = 1$ and using $b = 40$ bins.

can be concluded that adjusting the number of bins inherently involves a trade-off between accuracy and computational efficiency. Users are advised to carefully consider this balance, particularly in cases where ν is larger and the spatial correlation is strong.

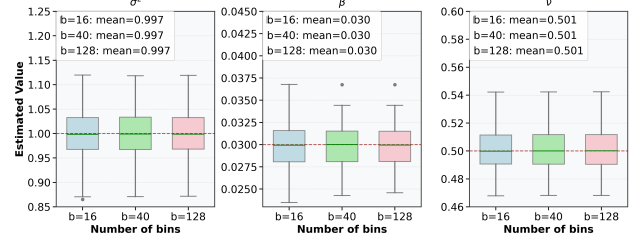
D. Accuracy Across Different Numbers of Bins

The numerical approximation accuracy of the BESSELK function, as defined in Equation (8), is significantly influenced by the discretization parameter b (number of bins). Following the previous discovery, we vary the number of bins ($b = 16, 40$, and 128) to assess its impact on the estimated parameters under weak, medium, and strong correlation levels when $\nu = 0.5$ and sample size $N = 51,076$, as illustrated in Figure 6. Additionally, we analyze the effect of the number of bins on the iteration count required for convergence across the three correlation levels, as shown in Figure 7.

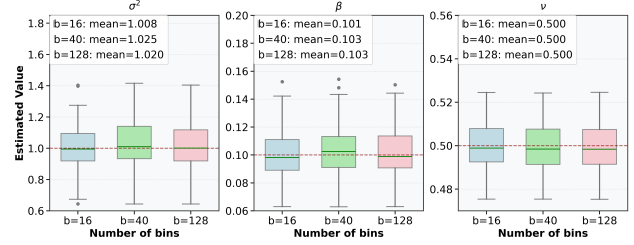
Although a smaller number of bins might introduce larger approximation errors given fixed integration bounds, this does not significantly affect the MLE procedure when ν is small. Through extensive numerical experiments (Figures 4, 5, 6, and 7), we have empirically demonstrated that parameter estimation remains robust even with a reduced number of bins, supporting our hypothesis that accurate estimation can be achieved without requiring fine-grained discretization, although discretization should be chosen based on the specific application and desired accuracy level according to the aforementioned analysis. Overall, our approach suggests a practical balance between computational efficiency and numerical accuracy in evaluating the Matérn covariance function.

E. Wind Speed Application

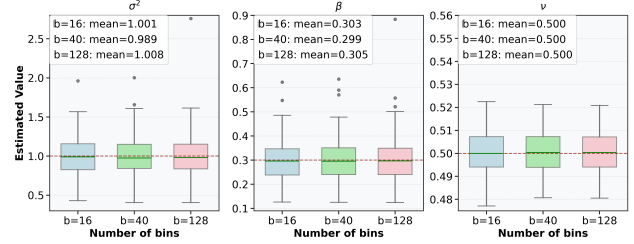
For real data analysis in this study, we use a wind speed dataset of 1M locations generated using the WRF-ARW model for the Arabian Peninsula. The model has a horizontal grid resolution of 5 km and 51 vertical levels, extending to a maximum altitude of 10 hPa. The dataset spans a geographical region from 20°E to 83°E



(a) Weak correlation $\beta = 0.03$.



(b) Medium correlation $\beta = 0.1$.



(c) Strong correlation $\beta = 0.3$.

Fig. 6: Parameter estimation over 100 replicas using the refined algorithm was evaluated for varying bin counts to estimate σ^2 , β , and ν across different correlation levels with problem size 51,076. Red dashed lines: true values of parameters.

longitude and 5°S to 36°N latitude, covering 37 years of daily records. Each file contains hourly wind speed measurements across 17 atmospheric layers. Our analysis specifically focused on wind speed data from September 1, 2017, at 00:00 AM, examining measurements at 10 meters above ground level (layer 0). To address the skewed distribution of wind speeds, we plot the square root of wind speed against longitude and latitude, as shown in Figure 8.

Starting with an initial dataset of 1M locations, we randomly sampled 160,000 locations for modeling and 25,000 for testing. To improve numerical stability, the location coordinates are preprocessed through normalization. Given the length and width of a squared region ℓ_1, ℓ_2 : (1) Compute the scaling factor $\ell := \max(\ell_1, \ell_2)$. (2) Rescale location (x_0, y_0) to $\{\{x_0 - \min(x_0)\}/\ell, \{y_0 - \min(y_0)\}/\ell\}$. This transformation maps the spatial coordinates into a unit square $[0, 1] \times [0, 1]$, which helps mitigate numerical issues in subsequent computations.

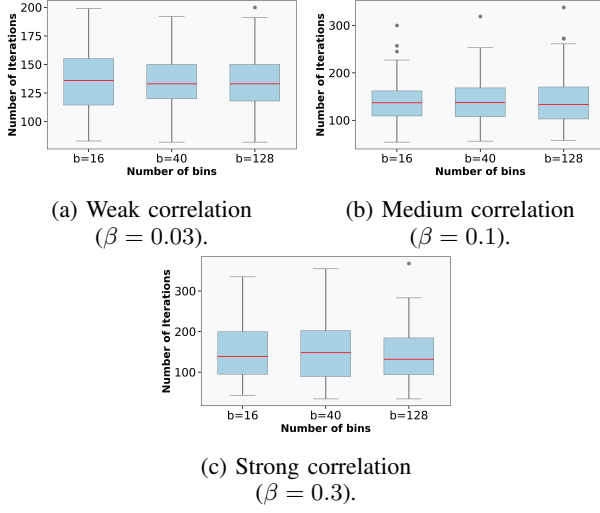


Fig. 7: The number of iterations for MLE optimization over 100 replicas using the refined algorithm was evaluated with $b = 16, 40$, and 128 bins across different correlation levels with problem size $51,076$, when $nu = 0.5$.

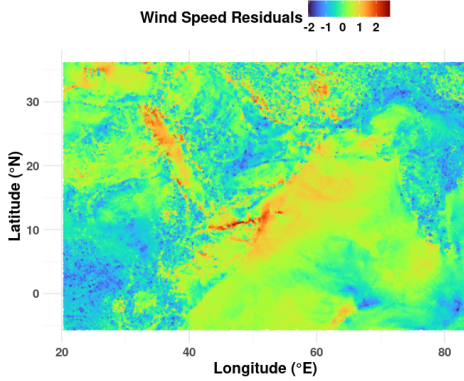


Fig. 8: Residuals of a wind speed dataset of 1M locations in the Middle East region.

Table I presents the estimated parameters and the final log-likelihood values obtained by the GSL and the refined algorithms. The experiments were carried out on a 40-core Intel Cascade Lake CPU with 383GB of memory and a 80GB single A100 GPU. Both methods estimated nearly identical parameters, achieving nearly the same maximum log-likelihood value (llh), and produced the same mean square prediction error (MSPE).

However, using the GSL library required 596.61 minutes of execution time, while the refined algorithm reduced this to 299.89 minutes. For fairness, only the matrix generation was performed either by the CPU (GSL) or the GPU (refined algorithm), while all other operations were executed on the GPU.

TABLE I: Comparison of GSL and refined algorithms on 160K random locations from the wind speed dataset.

	(σ^2, β, ν) estimates	llh value	MSPE	Time (min)
GSL	(2.505, 0.178, 0.426)	6984.660	0.037188	596.61
Refined	(2.510, 0.179, 0.426)	6984.598	0.037186	299.89

F. Performance Assessment

For single-node performance evaluation, we use a 40-core Intel Cascade Lake CPU to run the GSL version and generate an entire covariance matrix within the *ExaGeoStat* framework for varying numbers of locations. Additionally, we use 1 to 4 V100 (32 GB) or A100 (80 GB) GPUs on a single node to evaluate the implementation of the refined algorithm.

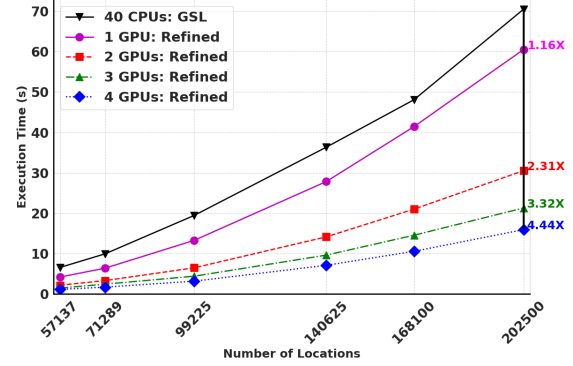


Fig. 9: Execution time for generating an $N \times N$ covariance matrix using GSL and the refined algorithm on a 40-core Intel Cascade Lake CPU and 1-4 NVIDIA V100 GPUs.

In the single-GPU (V100) configuration (Figure 9), tests on datasets with a number of locations ranging from 57,137 to 202,500 demonstrated that the refined algorithm achieved a 1.16X speedup compared to the CPU-only implementation. The performance improvements became increasingly pronounced with additional GPUs, reaching a 4.44X speedup with four GPUs. For NVIDIA A100 GPUs, the single-GPU configuration (Figure 10), tested on datasets with a number of locations ranging from 57,137 to 99,225, showed a 2.68X speedup over the CPU-only implementation. This improvement scaled significantly with additional GPUs, achieving a 12.62X speedup with four GPUs. These results highlight the effectiveness of GPU acceleration with the refined algorithm for matrix generation, with benefits becoming more pronounced when leveraging multiple GPUs and larger datasets.

Although the single-matrix generation time highlights the efficiency of the refined algorithm on GPUs compared to GSL, its impact becomes even more significant

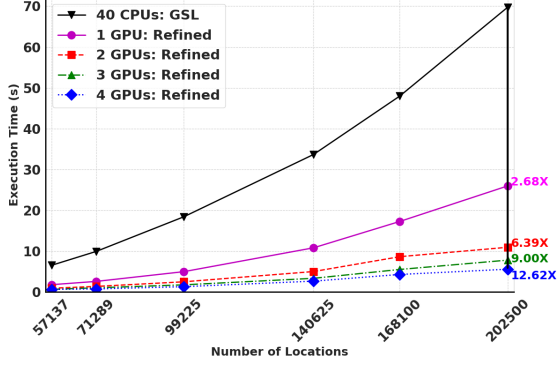


Fig. 10: Mean execution time (s) for generating an $N \times N$ covariance matrix using GSL and the refined algorithm on a 40-core Intel Cascade Lake CPU and 1-4 NVIDIA A100 GPUs.

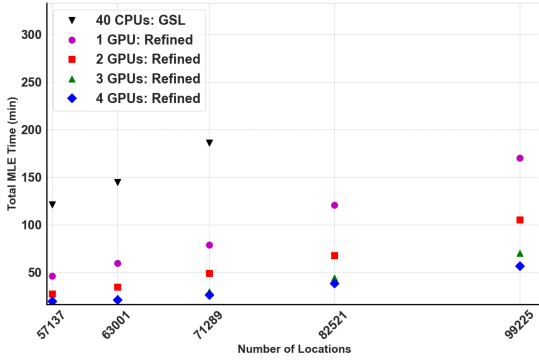


Fig. 11: Comparison of overall MLE execution time between the GSL library and the refined algorithm on a V100 GPU across various problem sizes. The reported time accounts for multiple iterations of the log-likelihood function, including matrix generation and all associated linear algebra operations within the MLE process [9].

in operations like MLE. During MLE, the matrix generation function is invoked multiple times while optimizing the log-likelihood function until convergence, amplifying the benefits of the refined algorithm. Figure 11 illustrates the performance of executing the full MLE process on five different problem sizes, comparing the GSL implementation with the refined algorithm using up to four A100 GPUs. Due to the long execution time of the GSL function, the full MLE process was not estimated for $N = 82,521$ and $N = 99,225$. For $N = 71,289$, the complete MLE process took 186.21 minutes with GSL (CPU) and 78.87, 49.36, 30.01, and 26.38 minutes using the refined algorithm on 1 GPU, 2 GPUs, 3 GPUs, and 4 GPUs, respectively.

Figure 12 shows the scalability of the refined algorithm across up to six nodes, each equipped with two V100/A100 GPUs. As available memory increases,

GPUs can handle larger problem sizes and the performance scales almost linearly with more GPUs.

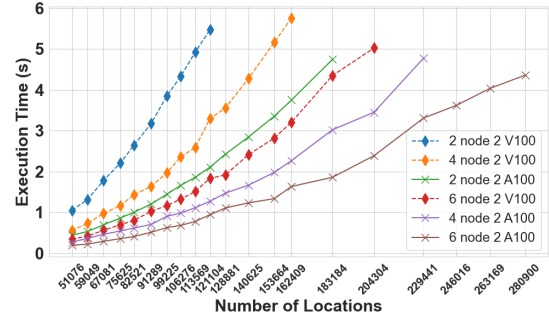


Fig. 12: Mean matrix generation time varying node counts and GPU architectures (V100 and A100).

VI. CONCLUSION

We present a highly efficient GPU-accelerated implementation of the modified Bessel function of the second kind (BESSELK) using CUDA, effectively addressing critical computational bottlenecks in Gaussian processes and various other scientific applications. This approach can also be extended to other modeling frameworks, including non-Gaussian processes, where BESSELK functions frequently arise. It integrates Temme's series expansion for small input values and a refined version of Takekawa's integral-based approach for larger values, ensuring accuracy and computational efficiency across a reasonable parameter space. Incorporating this optimized algorithm into the *ExaGeoStat* framework demonstrates significant performance improvements in generating covariance matrices for spatial data modeling. The GPU-based implementation achieved substantial speedups compared to traditional CPU-based methods while maintaining high numerical accuracy, validated through synthetic datasets and real-world climate data. These improvements are particularly notable in large-scale applications that require massive matrix computations. Our work improves the computational capabilities of BESSELK evaluations on GPUs and establishes the foundation for integrating such optimizations into other domains based on these functions. Future work includes extending the implementation to support the evaluation of derivatives of BESSELK, enabling gradient-based optimization techniques, as well as transitioning from single-threaded to a multi-threaded evaluation of $K_\nu(x)$ to further improve performance.

REFERENCES

- [1] Sharma, Kal Renganathan, "Damped wave conduction and relaxation in cylindrical and spherical coordinates," *Journal of Thermophysics and Heat Transfer*, vol. 21, no. 4, pp. 688–693, 2007.

- [2] Karamian, Asghar, Jazi, Bahram, and Najari, Samaneh, "The Role of Ordinary Bessel and Hankel Functions in Simulation of Plasma Valve Mechanism in a Loss-Free Metallic Cylindrical Waveguide," *Mathematics Interdisciplinary Research*, vol. 7, no. 4, pp. 343–355, 2022, University of Kashan.
- [3] Martin, Pablo, Rojas, Eduardo, Olivares, Jorge, and Sotomayor, Adrián, "Quasi-Rational Analytic Approximation for the Modified Bessel Function $I_1(x)$ with High Accuracy," *Symmetry*, vol. 13, no. 5, p. 741, 2021, MDPI.
- [4] Wang, Kesen, Abdulah, Sameh, Sun, Ying, and Genton, Marc G., "Which parameterization of the Matérn covariance function?," *Spatial Statistics*, vol. 58, p. 100787, 2023, Elsevier.
- [5] Takekawa, T., 2022. Fast parallel calculation of modified Bessel function of the second kind and its derivatives. *SoftwareX*, 17, p.100923.
- [6] Temme, N.M., 1975. On the numerical evaluation of the modified Bessel function of the third kind. *Journal of Computational Physics*, 19(3), pp.324–337.
- [7] Amos, Donald E., "Computation of modified Bessel functions and their ratios," *Mathematics of Computation*, vol. 28, no. 125, pp. 239–251, 1974.
- [8] F.W.J. Olver and L.C. Maximon, "Bessel Functions," in *NIST Handbook of Mathematical Functions*, Cambridge University Press, 2009, pp. 215–286.
- [9] S. Abdulah, H. Ltaief, Y. Sun, M. G. Genton, and D. E. Keyes, "ExaGeoStat: A high performance unified software for geostatistics on manycore systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 12, pp. 2771–2784, 2018.
- [10] Weisstein, Eric W., "Modified Bessel Function of the Second Kind," *MathWorld – A Wolfram Web Resource*, Wolfram Research, Inc., 2002. Available at: <https://mathworld.wolfram.com/>.
- [11] Dunster, T. Mark, "Bessel Functions of Purely Imaginary Order, with an Application to Second-Order Linear Differential Equations Having a Large Parameter," *SIAM Journal on Mathematical Analysis*, vol. 21, no. 4, 1990, pp. 995–1018, SIAM.
- [12] Zhukovsky, K. V., "Solving Evolutionary-Type Differential Equations and Physical Problems Using the Operator Method," *Theoretical and Mathematical Physics*, vol. 190, 2017, pp. 52–68, Springer.
- [13] Hetnarski, Richard B., and Eslami, M. Reza, "Heat Conduction Problems," *Thermal Stresses—Advanced Theory and Applications*, Springer, 2009, pp. 132–218.
- [14] Gómez-Correa, J. E., Balderas-Mata, S. E., Coello, V., Puente, N. P., Rogel-Salazar, J., and Chávez-Cerda, S., "On the Physics of Propagating Bessel Modes in Cylindrical Waveguides," *American Journal of Physics*, vol. 85, no. 5, 2017, pp. 341–345, AIP Publishing.
- [15] Pogány, Tibor K., "Bessel–Sampling Restoration of Stochastic Signals," *Acta Polytechnica Hungarica*, vol. 10, no. 7, 2013, pp. 7–19.
- [16] AlBahar, Areej, Kim, Inyoung, and Yue, Xiaowei, "A Robust Asymmetric Kernel Function for Bayesian Optimization, with Application to Image Defect Detection in Manufacturing Systems," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 4, 2021, pp. 3222–3233, IEEE.
- [17] Yin, Feng, Pan, Lishuo, Chen, Tianshi, Theodoridis, Sergios, Luo, Zhi-Quan Tom, and Zoubir, Abdelhak M., "Linear Multiple Low-Rank Kernel Based Stationary Gaussian Processes Regression for Time Series," *IEEE Transactions on Signal Processing*, vol. 68, 2020, pp. 5260–5275, IEEE.
- [18] Y.L. Luke, *Integrals of Bessel Functions*. Courier Corporation, 2014.
- [19] E. Grosswald, *Bessel Polynomials*, vol. 698. Springer, 1978.
- [20] M. Carley, "Numerical solution of the modified Bessel equation," *IMA Journal of Numerical Analysis*, vol. 33, no. 3, pp. 1048–1062, 2013.
- [21] J. Butcher, "Runge-Kutta methods," *Scholarpedia*, vol. 2, no. 9, p. 3147, 2007.
- [22] Gramacy, Robert B. *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. Chapman and Hall/CRC, 2020.
- [23] Nesi, Lucas Leandro, Arnaud Legrand, and Lucas Mello Schnorr. "Exploiting system-level heterogeneity to improve the performance of a geostatistics multi-phase task-based application." In *Proceedings of the 50th International Conference on Parallel Processing*, pp. 1–10, 2021.
- [24] The Chameleon project, "Chameleon: A dense linear algebra library for heterogeneous architectures," Dec. 2024. [Online]. Available: <https://project.inria.fr/chameleon/>
- [25] G. Bosilca, A. Bouteiller, A. Danalis, M. Faverge, A. Haidar, T. Herault, J. Kurzak, J. Langou, P. Lemarinier, H. Ltaief, *et al.*, "Flexible development of dense linear algebra algorithms on massively parallel architectures with DPLASMA," in *2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and PhD Forum*, IEEE, 2011, pp. 1432–1441.
- [26] C. Augonnet, S. Thibault, R. Namyst, and P. A. Wacrenier, "StarPU: a unified platform for task scheduling on heterogeneous multicore architectures," in *Euro-Par 2009 Parallel Processing: 15th International Euro-Par Conference, Delft, The Netherlands, August 25-28, 2009. Proceedings 15*, Springer, 2009, pp. 863–874.
- [27] R. Hoque, T. Herault, G. Bosilca, and J. Dongarra, "Dynamic task discovery in PARSEC: A data-flow task-based runtime," in *Proceedings of the 8th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems*, 2017, pp. 1–8.
- [28] S. Abdulah, H. Ltaief, Y. Sun, M. G. Genton, and D. E. Keyes, "Parallel approximation of the maximum likelihood estimation for the prediction of large-scale geostatistics simulations," in *2018 IEEE International Conference on Cluster Computing (CLUSTER)*, IEEE, 2018, pp. 98–108.
- [29] S. Abdulah, H. Ltaief, Y. Sun, M. G. Genton, and D. E. Keyes, "Geostatistical modeling and prediction using mixed precision tile Cholesky factorization," in *Proceedings of the 2019 IEEE 26th International Conference on High Performance Computing, Data, and Analytics (HiPC)*, 2019, pp. 152–162.
- [30] Q. Cao, S. Abdulah, H. Ltaief, M. G. Genton, D. Keyes, and G. Bosilca, "Reducing data motion and energy consumption of geospatial modeling applications using automated precision conversion," in *2023 IEEE International Conference on Cluster Computing (CLUSTER)*, IEEE, 2023, pp. 330–342.
- [31] X. Zhang, S. Abdulah, J. Cao, H. Ltaief, Y. Sun, M. G. Genton, and D. E. Keyes, "Parallel Approximations for High-Dimensional Multivariate Normal Probability Computation in Confidence Region Detection Applications," in *Proceedings of the 2024 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 265–276, IEEE, 2024.
- [32] J. B. Campbell, "On Temme's Algorithm for the Modified Bessel Function of the Third Kind," *ACM Transactions on Mathematical Software*, vol. 6, no. 4, pp. 581–586, Dec. 1980.
- [33] Watson, G. N. (1922). *A treatise on the theory of Bessel functions* (Vol. 2). The University Press.
- [34] Plesner, A., Sørensen, H.H.B. and Hauberg, S., 2024, May. Accurate Computation of the Logarithm of Modified Bessel Functions on GPUs. In *Proceedings of the 38th ACM International Conference on Supercomputing* (pp. 213–224).
- [35] Paul, C. Abbott. *Mathematica. In Revival: The Handbook of Software for Engineers and Scientists (1995)*, pages 926–962. CRC Press, 2018.
- [36] Walker, D.W., 2018. Morton ordering of 2D arrays for efficient access to hierarchical memory. *The International Journal of High Performance Computing Applications*, 32(1), pp.189–203.
- [37] Salvaña, Mary Lai O., Sameh Abdulah, Hatem Ltaief, Ying Sun, Marc G. Genton, and David E. Keyes. "Parallel space-time likelihood optimization for air pollution prediction on large-scale systems." *Proceedings of the Platform for Advanced Scientific Computing Conference*, 2022, pp. 1–11.
- [38] Sun, Ying, and Michael L. Stein. "Statistically and computationally efficient estimating equations for large spatial datasets." *Journal of Computational and Graphical Statistics*, vol. 25, no. 1, 2016, pp. 187–208. Taylor & Francis.