

Contribution ID: 619 Contribution code: **contribution ID 619**Type: **Poster**

## Accelerated Computation of a High Dimensional Kolmogorov-Smirnov Distance

Surrogate modeling and data-model convergence are important in any field utilizing probabilistic modeling, including High Energy Physics and Nuclear Physics. However, demonstrating that the model produces samples from the same underlying distribution as the true source can be problematic if the data is many-dimensional. The 1-D and multi-dimensional Kolmogorov-Smirnov test (ddKS) is a statistically powerful nonparametric test which can be implemented as a one- or two-sample test. We have developed three algorithms, one exact and two approximate, for the multi-dimensional Kolmogorov-Smirnov test proposed by Fasano. We apply ddKS to the comparison of photon distributions in the Belle II time-of-propagation detector using the collaboration's Geant4 simulation and our own neural network surrogate model. Additionally, we have derived an analytic form for the statistical significance of ddKS. Our approximations reduce the input time complexity from quadratic to log-linear (vdKS) and reduce the dimensional time complexity from exponential to linear (rdKS). The approximation methods maintain the statistical power of the exact method requiring tens of data points to indicate differences between most sampled distributions.

### Significance

Comparing multi-dimensional distributions efficiently is extremely important for applications that aim to replace expensive high-fidelity simulations with faster (possibly ML-based) methods. We have developed a metric that, unlike KL divergence, has the properties of a metric, and we present exact and approximate implementations.

### References

<https://arxiv.org/abs/2106.13706>

### Speaker time zone

Compatible with America

**Primary authors:** HAGEN, Alexander (Pacific Northwest National Laboratory); STRUBE, Jan (PNNL); HAINJE, Connor (PNNL); HAIDE, Isabel (Karlsruhe Institute of Technology); KAHN, James (Karlsruhe Institute of Technology (KIT)); Dr JACKSON, Shane (PNNL)

**Presenter:** Dr JACKSON, Shane (PNNL)

**Session Classification:** Posters: Crystal

**Track Classification:** Track 2: Data Analysis - Algorithms and Tools