Henry Wynn's Legacy
A Tribute Session
(a few words on ABCD)

Werner G. Müller

JOHANNES KEPLER UNIVERSITY LINZ

Drongen, June 15, 2026

mODa 14
Model-Oriented Data Analysis and Optimum Design

I first met Henry 1990 at mODa2

Photos are actually from mODa3 in 1992

ABCD: Approximate Bayesian Computation Design

- a convenient way to deal with difficult probability models and complicated model criteria;
- prerequisites: it is possible to sample from the probability model and a prior distribution for the model parameters;
- based on Simulation-based Bayesian optimal design (cf. Müller, 1999) and ABC (cf. Tavaré et al. 1997);
- independently proposed in Drovandi and Pettitt (2013) and Hainy, Müller and Wynn (2013).

Hainy, M., Müller, W.G., & Wynn, H.P. (2013). Approximate Bayesian Computation Design (ABCD), an introduction. In D. Ucinsky, A. C. Atkinson, and M. Patan, editors, *mODa 10 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, Springer International Publishing, 135–143.

Bayesian learning

Design criterion

$$U(\xi) = E_{y|\xi}[\Phi(\pi(\theta|y,\xi))] = \int_{y\in Y} \Phi(\pi(\theta|y,\xi))\pi(y|\xi)dy,$$

where $\Phi(\pi(\theta|y,\xi))$ denotes a general functional of the posterior distribution of θ.

Remark: The class of functionals such that for all $\pi(\theta)$ and $\pi(y|\theta,\xi)$

$$U(\xi) \geq \Phi(\pi(\theta))$$

is the class of convex functionals:

$$\Phi((1-\alpha)\pi_1 + \alpha\pi_2) \leq (1-\alpha)\Phi(\pi_1) + \alpha\Phi(\pi_2).$$

Estimation of $U(\xi)$

- If expression for $\Phi(\pi(\theta|y,\xi))$ or estimate $\hat{\Phi}(\pi(\theta|y,\xi))$ is available, perform Monte Carlo integration over the draws $\{y^{(i)},\ i=1,\ldots,G\}$:

$$\hat{U}(\xi) = \frac{1}{G}\sum_{i=1}^{G}\hat{\Phi}(\pi(\theta|y^{(i)},\xi)).$$

- Proceed by maximizing $\hat{U}(\xi)$ using any suitable stochastic optimization method.
- Estimation of $\Phi(\pi(\theta|y,\xi))$ if $\pi(\theta|y,\xi)$ is intractable $\rightarrow$ **ABC**.

ABC rejection sampling

Simple method to obtain approximate sample from posterior distribution $\pi(\theta|y)$:

- Draw parameter from the prior: $\theta \sim \pi(\theta)$.
- Draw variable y' from the probability model: $y' \sim \pi(y|\theta)$.
- Accept θ if $y' \in N_\epsilon(y)$.

ABC rejection sampler produces sample from approximate posterior

$$\pi(\theta|y' \in N_\epsilon(y)) \propto \int_{y' \in \mathcal{Y}} \mathbb{I}_{N_\epsilon(y)}(y')\pi(y'|\theta)\pi(\theta)dy'.$$

If y is high-dimensional, summary statistics $T(y)$ are used instead of y.

ABC algorithm for Optimal Design

1. choose a design ξ
2. draw $\{\theta^{(j)}, j = 1, \ldots, H\}$ from the prior distribution $\pi(\theta)$
3. for every sample parameter value $\theta^{(j)}$ draw from $\pi(y|\theta^{(j)}, \xi)$, hence producing a sample $S = \{(y'^{(j)}, \theta^{(j)}), j = 1, \ldots, H\}$ from the joint distribution of y and θ
4. produce a sample $O = \{y^{(i)}, i = 1, \ldots, G\}$ from the marginal distribution $\pi(y|\xi)$ by repeating steps 2 and 3
5. for every $y^{(i)} \in O$ collect the $\{(y'^{(j)}, \theta^{(j)})\} \in S$ for which $y'^{(j)}$ lies in a neighborhood $N_\epsilon(y^{(i)})$
6. approximate the criterion Φ using these neighborhood $\theta^{(j)}$ values
7. approximate $U(\xi)$ by Monte Carlo integration
8. repeat steps 1-7 for each ξ during an optimization procedure

Working on the paper, May 2025

The reward

Example: Quadratic regression

Probability model

$$\begin{aligned} y|\theta, X &\sim \mathcal{N}(X\theta, \sigma^2 I_N) \\ \theta &\sim \mathcal{N}(\theta_0, \sigma^2 R^{-1}) \end{aligned}$$

The rows of the design matrix X are quadratics in one factor, i.e. the i^{th} row is $(1, x_i, x_i^2)$.

All $x_i \in [-1, 1]$, $i = 1, \ldots, N$.

Criterion

- Criterion is Kullback-Leibler divergence/Shannon information:

$$\Phi(\pi(\theta|y,X)) = \int_{\theta\in\Theta} \log\left[\frac{\pi(\theta|y,X)}{\pi(\theta)}\right]\pi(\theta|y,X)d\theta$$

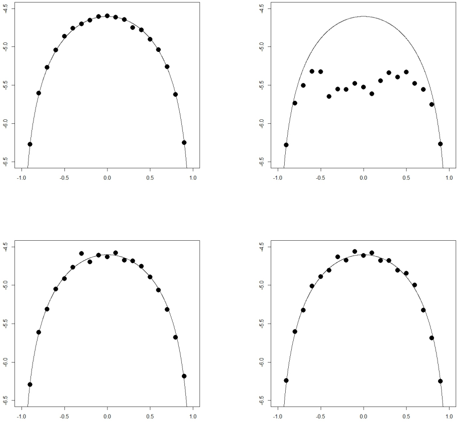
- For a linear regression model, maximizing $U(X) = \int_{y\in\mathcal{Y}} \Phi(\pi(\theta|y,X))p(y|X)dy$ is equivalent to maximizing the criterion for D_B optimality,

$$\Psi(X) = \det(M+R).$$

- Optimum continuous design puts equal weight of 1/3 on design points -1, 0, and 1.
- Prior information on points -1 and 1, i.e. $R = (1,-1,1)^T(1,-1,1) + (1,1,1)^T(1,1,1)$

$$\Rightarrow \boxed{\text{optimum at } x = 0}$$

Results



1: Top left: baseline

$G = 10^3$, $H = 10^5$

2: Top right: double ϵ

3: Bottom left: $H = 10^3$

4: Bottom right: $G = 10^2$

Embellishing ABCD

VIDEO!

Hainy, M., Müller, W. G., & Wynn, H.P. (2014) Learning functions and Approximate Bayesian Computation design: ABCD. *Entropy* 16 (8), 4353-4374. doi:10.3390/e16084353

London, summer of 2023

Gaussian Process Regression (GPR)

Observations to be generated from the model

$$y(x) = F^T(x)\,\theta + \varepsilon(x)\,; \quad x \in \mathcal{X},$$

where $\mathcal{X} = \{x_1, \ldots, x_N\}$ is a finite design space containing N points and θ is the unknown vector of parameters with $\dim(\theta) = p$.

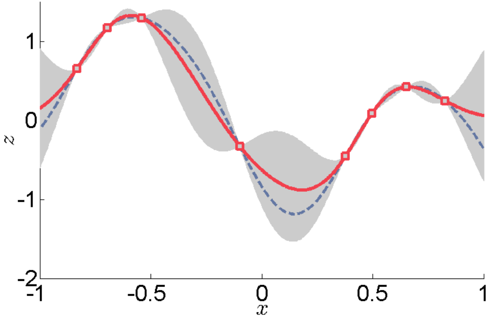
Further assumptions: $E\{\varepsilon(x)\} = 0$, and is correlated, defined through the $N \times N$ matrix C with elements

$$C_{ij} = \mathrm{Cov}\{\varepsilon(x_i), \varepsilon(x_j)\}\,; \qquad i, j = 1, \ldots, N.$$

The matrix C is assumed to be known and is typically parametrized via a covariance kernel (eg. Matérn).

We do not necessarily need to assume the errors to be Gaussian, but it is usually done, hence the name (also called universal kriging).

Prediction



A Typical Graph (from Wikipedia)

We observe the responses $y_1, y_2, \ldots, y_n$ at n-point design ξ and we want to predict the responses at all other $N - n$ unobserved locations $x_j \notin \xi$.

ABCD example: spatial prediction

- Compute predictive distribution of $y_{n+1}(x_{n+1})$ given $y(\xi) = (y_1(x_1), \ldots, y_n(x_n))$.
- Criterion of interest is the **maximum variance of the predictive distribution over the design space**

$$\Phi(y(\xi)) = \max_{x_{n+1}} \int (y_{n+1} - \mu_{y_{n+1}})^2 \pi(y_{n+1} | y(\xi), x_{n+1}) dy_{n+1}.$$

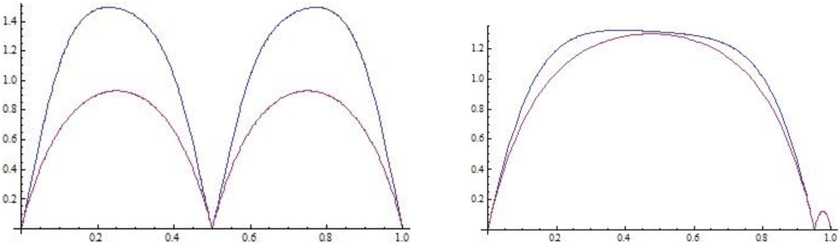
- The observations $(y_1(x_1), y_2(x_2), y_3(x_3), y_4(x_4))$ follow a (one-dimensional) Gaussian random field with mean zero, $x_i \in [0,1]$ and exponential correlation function

$$\rho(x_i, x_j) = \exp(-\theta \| x_i - x_j \|).$$

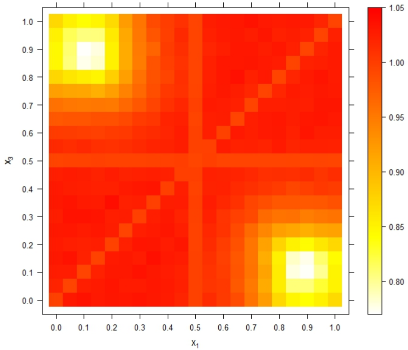
Spatial prediction design criterion: EK-optimality

From Müller et al. (2012), where

$$\min_{\xi} \max_{x} \left\{ \mathsf{Var}[\hat{y}(x)] + \mathsf{tr}\{M_\theta^{-1} \mathsf{Var}[\partial \hat{y}(x)/\partial\theta]\} \right\}$$



Criterion map

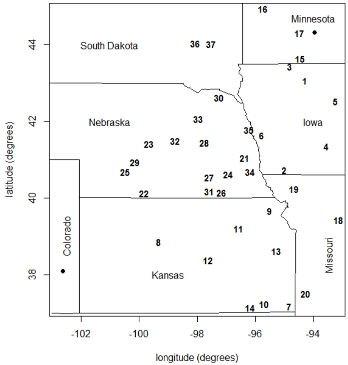


Map of approximated criterion values if $\theta = \log(100)$, $\epsilon = 0.1$, $H = 10^5$, and $G = 500$, $x_2 = 0.5$.

A final ABCD example: spatial extremes

- Goal: find optimal design $\boldsymbol{\xi}$ for spatial data following an extreme value distribution (maximum monthly temperature, maximum daily rainfall).
- Assume a max-stable process for the probability model of the data, e.g. Schlather model.
- The margins are assumed to be transformed to a unit-Fréchet distribution. Optimal design is performed with respect to the correlation parameters, $\boldsymbol{\vartheta}$, especially the range parameter (Whittle-Matérn).
- no likelihood function available for multivariate extreme value distribution for Schlather model if dimension > 2, but it is possible to simulate from the model.
- Optimality criterion: average posterior determinant of variance-covariance matrix of parameters over marginal distribution of data (*posterior precision utility*): $U(\boldsymbol{\xi}) = \int -\det[\mathsf{Var}(\boldsymbol{\vartheta}|y,\boldsymbol{\xi})]p(y|\boldsymbol{\xi})dy.$

Midwest temperature data



Data: maximum summer (June 1 – August 31) temperatures from 1895 to 2009 measured on 39 sites in the U.S. midwest region (115 observations per site), see Erhardt and Smith (2014).

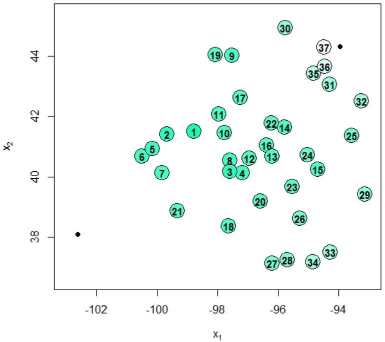
Design setting and summary statistics

- Consider only three-point designs.
- Points with greatest mutual distance are fixed, so number of designs is 37.
- Use tripletwise extremal coefficients as summary statistics (see Erhardt and Smith 2012).
- Each sample y'_r or $y'_{m,r}$, resp., consists of 1000 independent replicates of the extremal process.

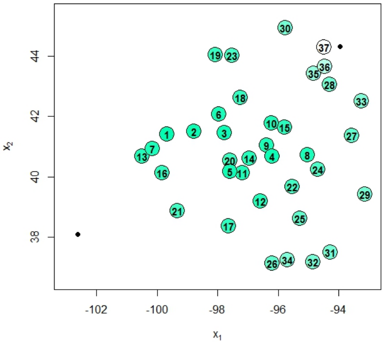
Simulation settings

- The sill parameter $c = 1$ and the smooth parameter $\kappa = 0.5$.
- The prior for the range parameter is $\lambda \sim U[2.5, 17.5]$.
- Simulation numbers:
  - $K = 2000$ for Monte Carlo integration
  - rejection sampler: keep 500 (0.01 %) of $R = 5,000,000$ particles.
  - importance weights: $R = 2000$ particles, $M = 4000$

Criterion rankings for the precision utility



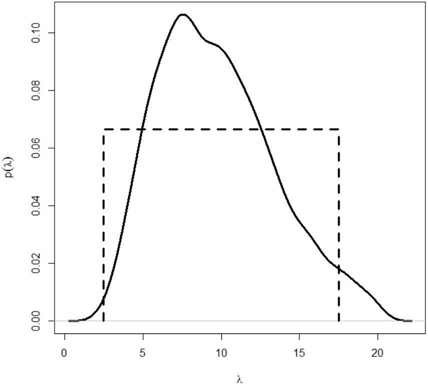
ABC rejection



importance weight updates

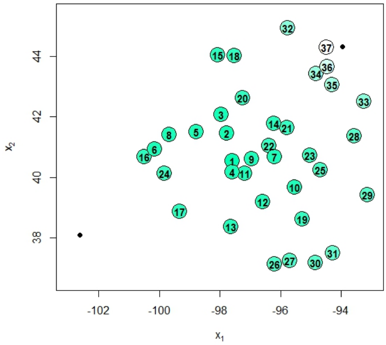
ABC estimation results for range parameter

An ABC sampler was run for $R = 5 \cdot 10^6$ iterations, $N = 2000$ draws were kept (0.02 %).

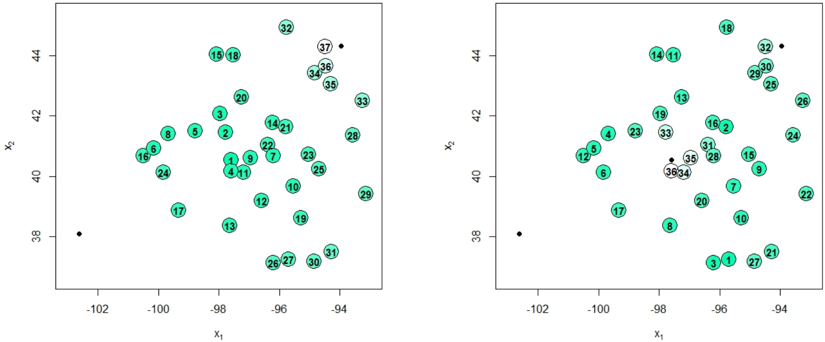


Criterion rankings if ABC sample is used as prior

Results for ABC rejection algorithm if ABC sample is used as prior; other simulation settings as before



Design augmentation $(3 \to 4)$



Design augmentation (4 → 5)