# Bayesian Ranking of Treatments for Experiment Evaluation and Feedback Intervention


Miguel R. Pebes-Trujillo[1,3], Itamar Shenhar[2,3], Aravind Harikumar[1,3], Ittai Herrmann[2,3], Menachem Moshelion[2,3], Kee Woei Ng[1,3], Matan Gavish[2,3]

[1]Nanyang Technological University, [2] Hebrew University of Jerusalem, [3] Singapore-HUJ Alliance for Research and Enterprise



BayesComp 2025 Singapore 16 - 20June


## Abstract

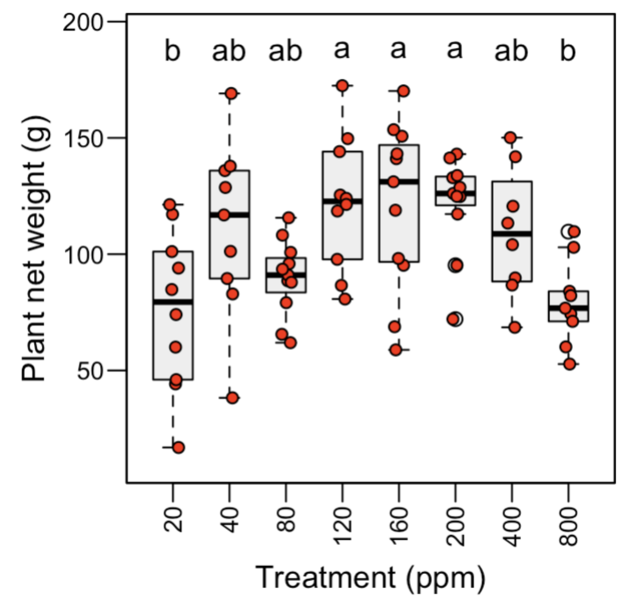

Experimental research often involves applying treatments to individuals to identify the one that best achieves the desired characteristics. **Our method finds the optimal treatment** with limited data and conflicting trait goals. Applied diachronically, it **enables continuous trait optimization**.


## Example 1: Experiment Evaluation

We analyze harvest plant weight measurements from a precision agriculture experiment where treatments varied in the supplied nitrogen concentration: 20, 40, 80 120, 160, 200, 400, or 800 mg $L^{-1}$ or parts per million (ppm). Which one maximizes yield (∼weight)?

| **Typical analysis (inputs data)** | **Our method (inputs data + optimization criteria)** |
|---|---|

**Net Weight Distribution**



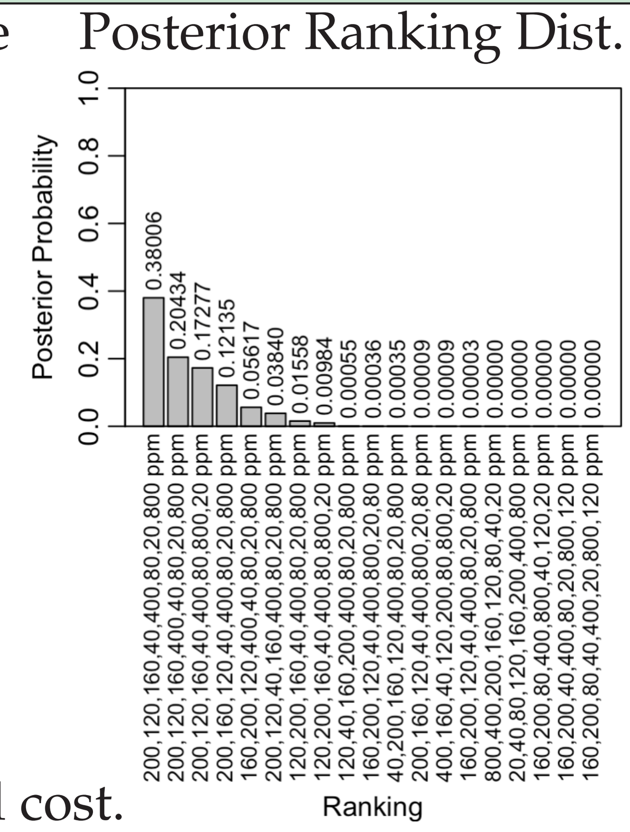

**Tukey HSD Test P-values**

| Treatment (ppm) | 20 | 40 | 80 | 120 | 160 | 200 | 400 | 800 |
|---|---|---|---|---|---|---|---|---|
| 800 | 1 | 0.27 | 0.99 | 0.03 | 0.04 | 0.03 | 0.38 | 1 |
| 400 | 0.23 | 1 | 0.82 | 0.98 | 0.99 | 0.98 | 1 | 0.38 |
| 200 | 0.01 | 0.99 | 0.16 | 1 | 1 | 1 | 0.98 | 0.03 |
| 160 | 0.01 | 0.99 | 0.19 | 1 | 1 | 1 | 0.99 | 0.04 |
| 120 | 0.01 | 0.99 | 0.18 | 1 | 1 | 1 | 0.98 | 0.03 |
| 80 | 0.95 | 0.71 | 1 | 0.18 | 0.19 | 0.16 | 0.82 | 0.99 |
| 40 | 0.14 | 1 | 0.71 | 0.99 | 0.99 | 0.99 | 1 | 0.27 |
| 20 | 1 | 0.14 | 0.95 | 0.01 | 0.01 | 0.01 | 0.23 | 1 |

**Ranking Point Estimate**

1) **200 ppm**
2) 120 ppm
3) 160 ppm
4) 40 ppm
5) 400 ppm
6) 80 ppm
7) 20 ppm
8) 800 ppm

**Posterior Ranking Dist.**




**Prevaling Probs (row over col)**

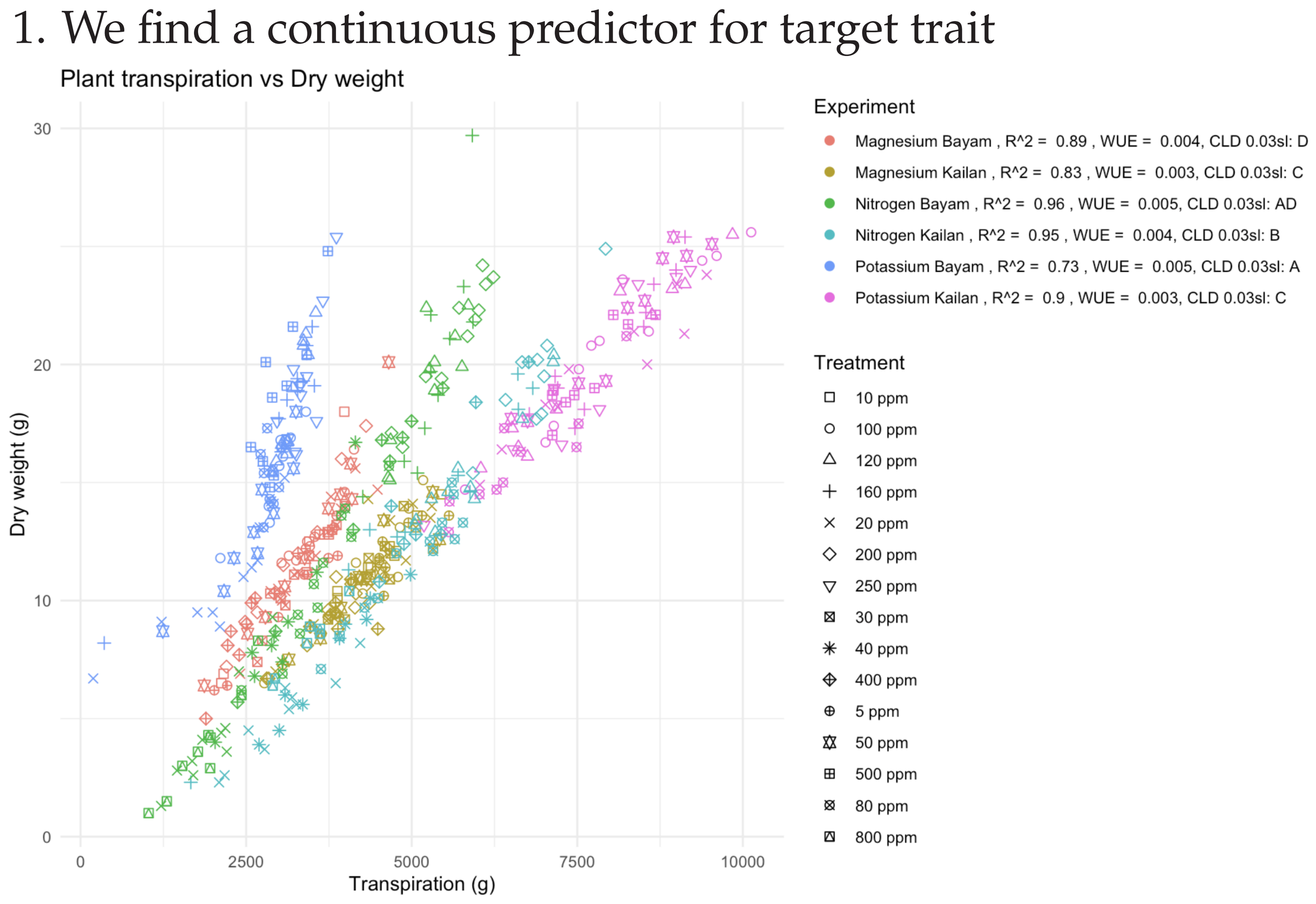
| Treatment (ppm) | 20 | 40 | 80 | 120 | 160 | 200 | 400 | 800 |
|---|---|---|---|---|---|---|---|---|
| 800 | 0.46 | 0.18 | 0.37 | 0.13 | 0.14 | 0.11 | 0.2 | 0.5 |
| 400 | 0.77 | 0.46 | 0.69 | 0.37 | 0.38 | 0.32 | 0.5 | 0.8 |
| 200 | 0.88 | 0.65 | 0.83 | 0.56 | 0.57 | 0.5 | 0.68 | 0.89 |
| 160 | 0.84 | 0.58 | 0.78 | 0.49 | 0.5 | 0.43 | 0.62 | 0.86 |
| 120 | 0.85 | 0.59 | 0.79 | 0.5 | 0.51 | 0.44 | 0.63 | 0.87 |
| 80 | 0.6 | 0.28 | 0.5 | 0.21 | 0.22 | 0.17 | 0.31 | 0.63 |
| 40 | 0.79 | 0.5 | 0.72 | 0.41 | 0.42 | 0.35 | 0.54 | 0.82 |
| 20 | 0.5 | 0.21 | 0.4 | 0.15 | 0.16 | 0.12 | 0.23 | 0.54 |

In practice, we want to achieve multiple (possibly conflicting) traits. For example, maximizing yield and nutritional content while minimazing environmental impact and cost.

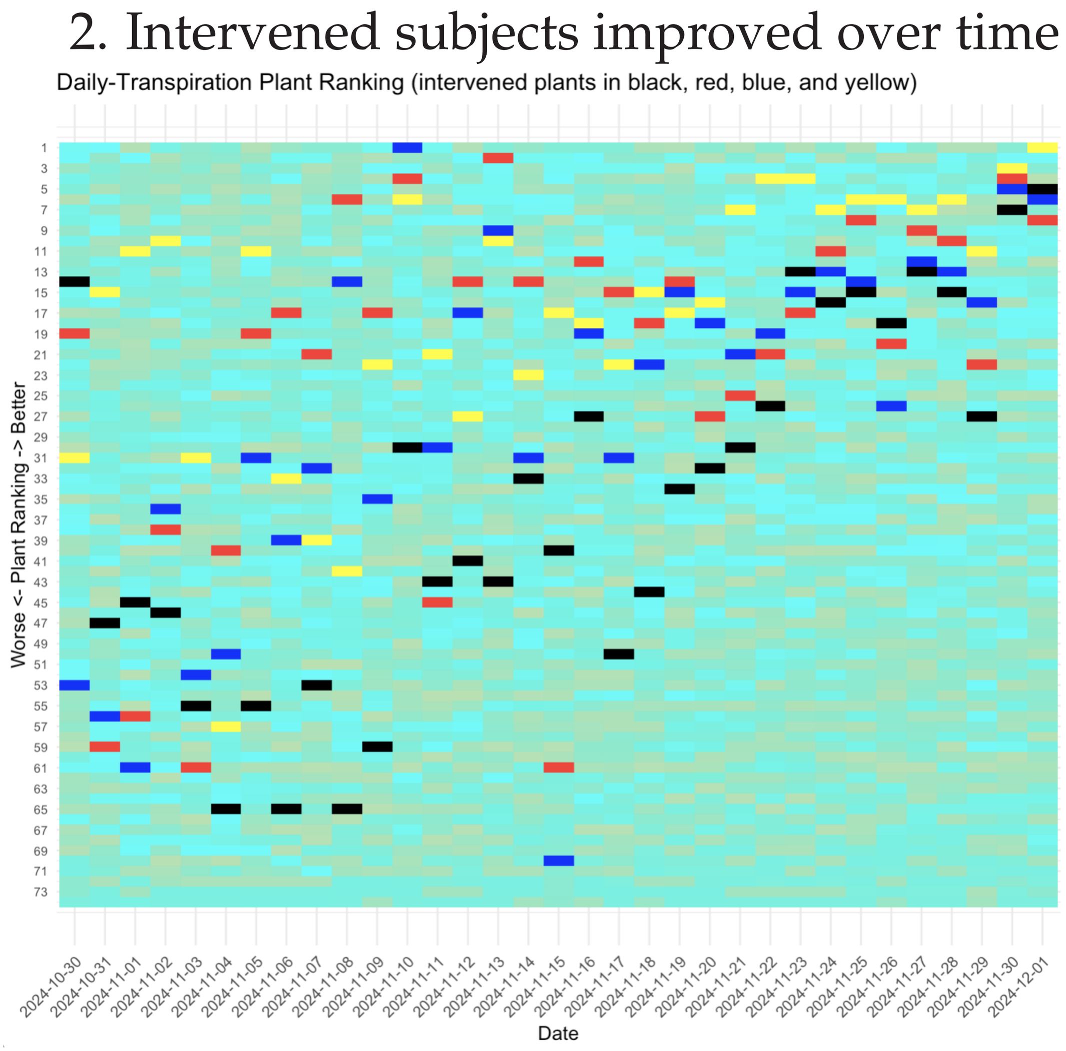
## Example 2: Feedback Intervention (Bayesian online learning)

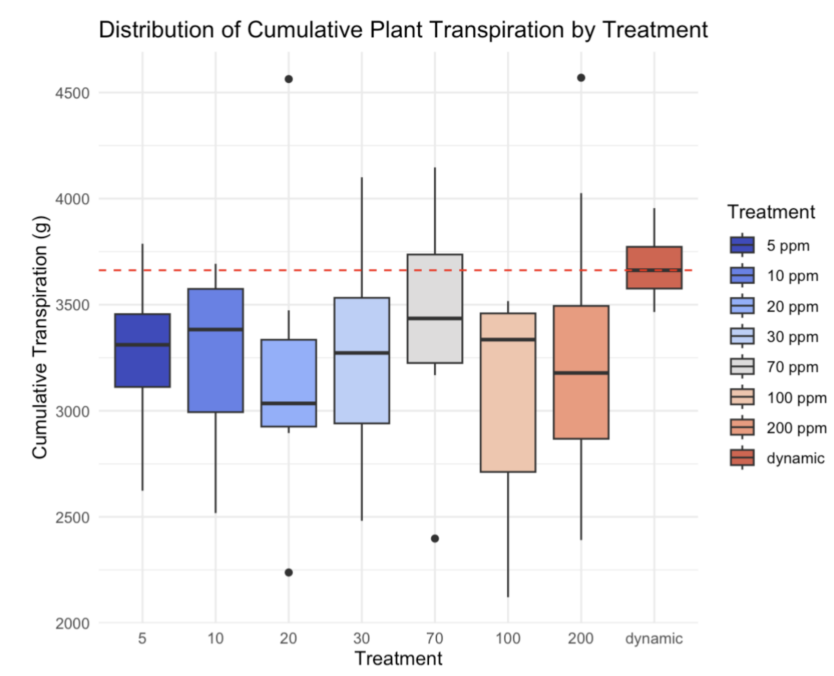
Consider the problem of **yield maximization**. We conducted a experiment where 70 Kailan (Chinese broccoli) plants where subjected to 7 treatments (varying in Phosphorus concentration). 4 plants where subjected to a fertigation regime where **our Bayesian model suggests the treatment these plants should receive each day**.

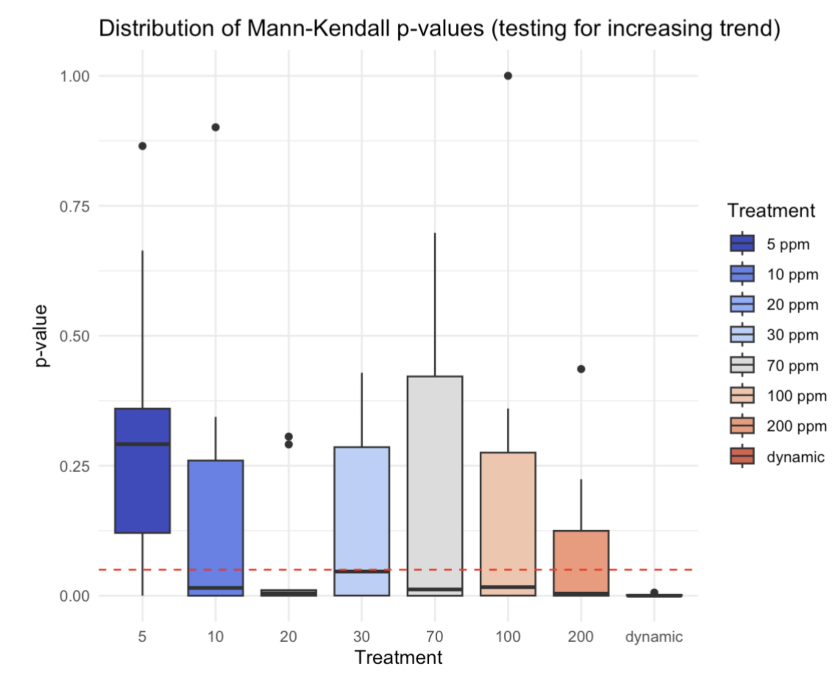
1. We find a continuous predictor for target trait




2. Intervened subjects improved over time




## Key Ideas on Characterization and Modeling

**Problem**: Determine which of the $K$ treatments is the most effective in producing a desired set of $M$ ideal features $\boldsymbol{X}^* = \{X_1^*, \ldots, X_M^*\} \in \mathbb{R}^M$.
**Data**: $\boldsymbol{X}_k = [X_{kij}] \in \mathbb{R}^{n_k \times M}; k = 1, \ldots, K; i = 1, \ldots, n_k; j = 1, \ldots, M.$
**Deviation RVs:** $D_{kij} = |X_j^* - X_{kij}|$ (deviation symmetry assumption)
**Treatment comparison RVs**: For two treatments $r$ and $s$

$$Y_{jrs} = \sum_{a=1}^{n_r} \sum_{b=1}^{n_s} I(D_{raj} < D_{sbj}).$$

Then $\boldsymbol{Y} = [Y_{jrs}] \in (\{0\} \cup \mathbb{Z}^+)^{M \times K \times K}$ is the collection of $Y_{jrs}$ for all features $j$ and all treatment pairs $(r, s)$; and $\boldsymbol{y}$ = observed realizations.
**Treatment Performance Characterization**: via em dominance indexes

$$\boldsymbol{d} = d_1, \ldots, d_K \mathbb{R}^K$$

**Reconciling multiple features**: Each treatment has $M$ sub-indexes $d_{k1}, \ldots, d_{kM} \in [0, u]^M$, s.t. $d_k = d_{k1} + \ldots + d_{kM}$, and with $\sum_{j=1}^M w_{kj} = 1$

$$d_{kj} = d_k w_{kj}.$$

Denote weights $\boldsymbol{w}_k = [w_{k1}, \ldots, w_{kM}]$, and $\boldsymbol{w} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K]^\top$.
**The Bradley-Terry Model for Experiments (BTME)** Consider the BT model with $\phi : \mathbb{R} \to (0, 1)$, e.g. sigmoid $\phi(x) = (1 + e^{-x})^{-1}$, for $\boldsymbol{y}$:

$$\begin{aligned} p(\boldsymbol{y} \mid \boldsymbol{n}, \boldsymbol{d}, \boldsymbol{w}) &= \prod_{j=1}^{M} \prod_{r=1}^{K} \prod_{s=r+1}^{K} \binom{n_r \times n_s}{y_{jrs}} \times (\phi(d_r w_{rj} - d_s w_{sj}))^{y_{jrs}} \\ &\times \quad (1 - \phi(d_r w_{rj} - d_s w_{sj}))^{n_r \times n_s - y_{jrs}}. \end{aligned}$$

## Key Steps on Inference

We use vanilla M-H to estimate $(\boldsymbol{d}, \boldsymbol{w})$, which, jointly, answer our question and enable uncertainty quantification. Sample parameterization:

$$\begin{aligned} d_1, \ldots, d_K &\overset{iid}{\sim} \text{Uniform}[0, u \times M] \quad (k = 1, \ldots, K) \\ w_{k1}, \ldots, w_{kM} &\sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_K) \\ \pi_{rs} &= \text{Sigmoid}(d_r w_{rj} - d_s w_{sj}) \quad (j = 1, \ldots, M; r < s) \\ y_{jrs} \mid n_r, n_s, \pi_{rs} &\overset{ind}{\sim} \text{Binomial}(n_r \times n_s, \pi_{rs}) \quad (j = 1, \ldots, M; r < s); \end{aligned}$$

Proposals: Trucated-Normal$(0, u \times M)$ for $d$s, and we randomly walk on the $(M - 1)$-Simplex for $\boldsymbol{w}$s. Finally, we estimate vector $\boldsymbol{d}$ and matrix $\boldsymbol{w}$.

## References (Full list in paper QR)


[1] Miguel R. Pebes-Trujillo, Itamar Shenhar, Aravind Harikumar, Ittai Herrmann, Menachem Moshelion, Kee Woei Ng, and Matan Gavish. Ranking of multi-response experiment treatments. *arXiv*. 2024.


## Acknowledgement


Research supported by the NRF Singapore, under its Campus CREATE and program SHARE REQ0246254 (NRF2020-THE003-0008).