

Uncertainty-aware breeding decisions: MCMC-based optimum contribution selection increases breeding decision robustness

Jon Ahlinder ¹, Patrik Waldmann ^{2,*}

¹Department of Tree Breeding, The Forest Research Institute of Sweden (Skogforsk), Tomterna 1, Sävar, SE 91833, Sweden

²Mathematical Sciences, University of Oulu, Paavo Havaksen tie 3, Oulu, FI 90570, Finland

*Corresponding author: Mathematical Sciences, University of Oulu, Paavo Havaksen tie 3, Oulu, FI, 90570, Finland. Email: patrik.waldmann@oulu.fi

Optimum contribution selection (OCS) balances genetic gain and inbreeding by optimizing parental contributions to the next generation, but current implementations rely on point estimates of breeding values that discard the uncertainty inherent in genetic evaluations. We introduce CVaR-OCS, a novel formulation that incorporates the full posterior distribution of estimated breeding values (EBVs) directly into the OCS objective via Conditional Value at Risk (CVaR) (a coherent risk measure from financial portfolio theory) allowing a single optimization to simultaneously maximize expected genetic gain and protect against worst-case outcomes driven by EBV uncertainty. We evaluate CVaR-OCS on a simulated multi-generation genomic selection dataset with known true breeding values (QTL-MAS 2010; $n = 3,226$), enabling direct comparison to an oracle solution, and on Norway spruce (*Picea abies* $n = 5,525$) and Loblolly pine (*Pinus taeda* $n = 926$) forest tree breeding progeny trials, with consistent tail-gain protection observed across all three datasets. On the simulated dataset, point-estimate MAP-OCS recovered only 78% of the genetic gain achieved by an oracle solution with access to true breeding values, illustrating the cost of ignoring predictive uncertainty; CVaR-OCS attained comparable expected gain while improving tail-gain security ($\text{CVaR}_{95} + 0.69\%$) and recovering an additional oracle-optimal individual. In Norway spruce, the recommended CVaR-OCS operating point improved tail-gain security by 6.60% and broadened the selection base from 145 to 159 individuals at a genetic gain cost of only 0.70%. Complementary MCMC-based robustness scores revealed that 25 MAP-OCS selections in Norway spruce were unstable across the posterior distribution; post-hoc exclusion of these individuals failed to improve tail-gain security, motivating the principled CVaR-OCS approach. CVaR-OCS provides breeders with a principled, computationally efficient tool for uncertainty-aware selection decisions, and multi-generation simulation studies are needed to fully characterize its long-term effects on genetic gain trajectories and inbreeding accumulation.

Keywords: conditional value at risk; selection optimization; estimated breeding value uncertainty; Bayesian genetic evaluation; Norway spruce; posterior uncertainty

Introduction

Maximizing genetic merit while restricting relatedness in a breeding population is a major challenge in many breeding programs (Goddard 1994; Lynch and Walsh 1998). If relatedness is not controlled, elevated levels of inbreeding will cause fitness depression and reduce the performance of the offspring in the population (Lynch and Walsh 1998; Charlesworth and Willis 2009). Typically, the genetic load will increase, that is, the presence of deleterious alleles in the population, leading to a higher incidence of genetic disorders (Derks and Steensma 2021; Tao et al. 2025). Therefore, much effort has been devoted to developing methods that optimize the balance between gain and relatedness of the selected set of individuals and their contribution to the next generation of breeding or deployment population (Goddard 1994; Mullin and Belotti 2015; Hamilton 2020; Cowling et al. 2023; Tchounke et al. 2023).

One of the most popular approaches, the optimum contribution selection (OCS), optimizes a linear objective function with

quadratic constraints on relatedness (inbreeding) (Meuwissen 1997). To solve the optimization problem, the original approach uses Lagrangian multipliers in an iterative fashion where non-selected individuals are excluded one by one. Since then, many efforts have been made to further improve the original approach by introducing enhanced optimization techniques such as semidefinite programming (SDP) (Pong-Wong and Woolliams 2007), the branch-and-bound algorithm (Mullin and Belotti 2015), second-order cone programming (SOCP) (Yamashita et al. 2018), matrix algebra computational tricks (Hinrichs et al. 2006), comparing the effect of kinship matrices on genetic gain, inbreeding and diversity (Gautason et al. 2023), including mate allocation optimization (Waldmann 2025) and various alternative formulations of the optimization problem to fit a range of different applications (Hallander and Waldmann 2009; Clark et al. 2013; Hamilton 2020; Hjortø et al. 2022; Fogg et al. 2024). Computational efficiency is of particular importance for the use of OCS in large-scale national and international breeding

programs, where pedigrees can comprise millions of individuals, such as the Interbull international cattle genetic evaluation system (Weigel et al. 2001), the Breedbase platform used in plant breeding (Hershberger et al. 2021), and TreePlan in forest tree breeding (McRae et al. 2003), making single-solve optimization approaches preferable to methods that require repeated optimization in many scenarios.

The efficiency of OCS relies heavily on the accuracy of the estimated breeding values (EBVs) used to calculate the contributions of breeding candidates. Typically, EBVs are obtained as point estimates that are plugged directly into OCS, so that uncertainty in those estimates is not accounted for in the optimization and few studies have investigated this aspect. Fogg et al. (2024) introduced a robust OCS framework drawing on concepts from robust optimization, in which each candidate's EBV uncertainty is represented by its standard error (SE), obtainable from frequentist inference or as a summary statistic from Bayesian methods. The robust OCS penalizes candidates whose high EBVs are accompanied by large SEs, preferring individuals with more reliable estimates even at some cost to expected gain; as the risk-aversion parameter approaches its maximum, this framework reduces to a minimax worst-case formulation that guards against the single most adverse realization of EBV uncertainty. Pocrnic et al. (2023) proposed a complementary probabilistic approach, applying OCS across posterior draws of EBVs from a Bayesian MCMC analysis of a simulated breeding population. Their results revealed considerable bias in the optimized contributions of standard OCS relative to the probabilistic approach, with the latter producing solutions closer to the true optimum, demonstrating that ignoring the uncertainty of EBVs in the selection process can lead to suboptimal decisions, even when the point estimates themselves are unbiased. However, the approach of Pocrnic et al. (2023) was evaluated on a small simulated pedigree, leaving open whether ignoring EBV uncertainty systematically compromises genetic progress in operational breeding populations (Bijma 2012), and whether a single-solve uncertainty-aware formulation, rather than repeated per-draw optimization, can achieve equivalent or superior protection at a fraction of the computational cost. We refer to the point-estimate approach throughout as MAP-OCS (maximum a posteriori OCS).

Both approaches above characterize EBV uncertainty marginally: robust OCS based on SE (Fogg et al. 2024) and per-draw posterior sampling (Pocrnic et al. 2023) treat each candidate's uncertainty independently, leaving unexploited a critical property of Bayesian MCMC genetic evaluation, i.e. the joint structure of the posterior (Sorensen and Gianola 2002). The full posterior $p(\mathbf{g} | \mathbf{y})$ captures not only the marginal variance of each individual's breeding value but also the pairwise prediction error covariances (PEVs) arising from shared relatives, overlapping phenotypic information and genomic relationships (Henderson 1975). By sampling directly from this joint posterior at each MCMC iteration, the full PEV covariance structure propagates implicitly into the contribution optimization. Two candidates with positively correlated EBV uncertainty are jointly riskier to select together than two candidates with the same marginal SE whose uncertainties are independent or negatively correlated, a distinction that becomes especially consequential in populations with strong family structure or high genomic relatedness among elite candidates.

Exploiting this joint structure calls for a framework designed for decision-making under distributional uncertainty: portfolio optimization, in which the goal is to identify an allocation,

consisting of a vector of parental contributions, that performs well not just under a single assumed scenario but robustly across the full distribution of possible outcomes (Markowitz 1952). In financial portfolio theory, the standard tool for quantifying and minimizing downside risk is the Conditional Value at Risk (CVaR), also known as Expected Shortfall (Rockafellar and Uryasev 2000, 2002). CVaR at confidence level α measures the expected loss in the worst $(1 - \alpha)$ fraction of scenarios. In the OCS context, this corresponds to the expected genetic gain when breeding values fall in their least favorable tail across MCMC posterior draws. Crucially, CVaR is a coherent risk measure (Artzner et al. 1999), meaning it satisfies subadditivity. Hence, diversification across individuals can only reduce, never increase, tail risk. This makes CVaR particularly well suited to OCS, where contribution diversification is already the central mechanism for managing inbreeding. CVaR-OCS directly addresses the limitations of both prior approaches. Unlike the SE-based robust OCS of Fogg et al. (2024), which treats each candidate's uncertainty independently and marginally, CVaR-OCS propagates the full joint posterior uncertainty into the optimization through the MCMC scenario set, and subsumes the Fogg et al. (2024) framework as a limiting case when the confidence level α approaches one, and in contrast to the per-draw approach of Pocrnic et al. (2023), CVaR-OCS requires solving only a single quadratic program rather than one per posterior sample, making it computationally practical for operational breeding programs with large pedigrees.

The aim of our study is threefold. First, we introduce CVaR-OCS, a novel formulation of optimum contribution selection that directly incorporates the joint posterior distribution of EBVs into the optimization objective via Conditional Value at Risk (CVaR). CVaR-OCS solves a single quadratic program that simultaneously maximizes expected genetic gain and minimizes tail risk arising from EBV uncertainty, providing a principled and computationally efficient alternative to point-estimate OCS that requires no repeated optimization across MCMC samples. Second, we use Monte Carlo sampling across MCMC iterations (MCMC-OCS) as a diagnostic tool to characterize how EBV uncertainty propagates into contribution decisions, quantifying the instability of MAP-OCS selections through overlap metrics, selection frequency analysis, and individual robustness scores. These diagnostics motivate CVaR-OCS and provide complementary insight into the sources of selection instability. Third, we evaluate CVaR-OCS and its MCMC-OCS diagnostics on a multi-generation genomic selection simulation using the QTLMAS 2010 benchmark dataset (Szydlowski and Paczyńska 2011) with known true breeding values, enabling direct comparison to an oracle solution and assessment of decision quality under controlled EBV uncertainty and on an operational Norway spruce (*Picea abies*) and Loblolly pine (*Pinus taeda*) forest tree breeding pedigrees.

Materials and methods

Optimum contribution selection

Let n denote the number of individuals in the recruitment or candidate population and let $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ be the matrix of pairwise genetic relationships among candidates, inferred from pedigree information, marker data, or a combination of both (see Case studies below for the specific matrix used in each analysis). For optimum contribution selection (OCS), the optimization problem

can be formulated as (Pong-Wong and Woolliams 2007; Ahlinder et al. 2013; Mullin and Belotti 2015; Yamashita et al. 2018):

$$\text{maximize}_{\mathbf{c}} \quad \mathbf{c}^T \mathbf{g} \quad (1)$$

$$\text{subject to} \quad \frac{1}{2} \mathbf{c}^T \mathbf{\Sigma} \mathbf{c} \leq \Theta, \quad (2)$$

$$\mathbf{c}^T \mathbf{s} = 0.5, \quad (3)$$

$$\mathbf{c}^T \mathbf{d} = 0.5, \quad (4)$$

$$\mathbf{c}^T \mathbf{1} = 1, \quad (5)$$

$$\mathbf{c} \leq \mathbf{m}, \quad (6)$$

$$\mathbf{c} \geq \mathbf{0}, \quad (7)$$

where $\mathbf{c} \in \mathbb{R}^n$ is the vector of individual contributions (the decision variable), $\mathbf{g} \in \mathbb{R}^n$ is the vector of estimated breeding values (EBVs) of the selection candidates, Θ is the upper bound on group coancestry of the selected cohort, $\mathbf{m}$ is a vector of upper bounds on the contribution of any single individual arising from practical limitations, and $\mathbf{s}$, $\mathbf{d}$ are indicator vectors for sires and dams, respectively. Note that the gender constraints (3)–(4) are included for dioecious species and removed when the species is monoicous, in which case the total-contribution constraint (5) ensures $\mathbf{c}^T \mathbf{1} = 1$. The objective function (1) is linear in $\mathbf{c}$ with a quadratic constraint on average relatedness among the selected candidates, so the problem (1)–(7) is a quadratic programme (QP). Here, we used the COSMO solver (Garstka et al. 2021) available in the JuMP modelling language for Julia (Dunning et al. 2017). This solver handles convex optimization problems with quadratic objective functions and conic constraints using an operator splitting algorithm based on ADMM (Boyd et al. 2010). Please consult [Supplementary Table S5](#) for further details on the tested optimization methods and [Supplementary Table S9](#) for computational performance on the analyzed pedigrees.

Bayesian mixed model and MCMC inference

Following [Sorensen and Gianola \(2002\)](#), we used the following multi-trait mixed effects model to infer breeding values:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}, \quad (8)$$

where $\mathbf{y}$ is the vector containing each of m traits represented sequentially for n individuals in a single column vector of size $nm \times 1$; $\mathbf{X}$ and $\mathbf{Z}$ are the fixed and random-effect design matrices of sizes $nm \times pm$ and $nm \times nm$, respectively, associating predictors with the corresponding phenotypic records for p levels of fixed-effect predictors. Random polygenic and residual effects are assumed to follow, for two traits, one categorical and one continuous:

$$\begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix} \middle| \mathbf{\Sigma}, \mathbf{G}_0 \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \mathbf{G}_0 \otimes \mathbf{\Sigma} \right), \quad (9)$$

where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the vectors of breeding values (polygenic scores) for the continuous trait (subscript 1) and the categorical trait (subscript 2), respectively. In addition, $\mathbf{\Sigma}$ is the $n \times n$ genetic relationship matrix of the candidates (pedigree $\mathbf{A}$, genomic $\mathbf{G}$, or blended $\mathbf{H}$ as specified per case study) and $\mathbf{G}_0$ is the 2×2 genetic covariance matrix between traits, defined as

$$\mathbf{G}_0 = \begin{pmatrix} \sigma_{a_1}^2 & \sigma_{a_1, a_2} \\ \sigma_{a_1, a_2} & \sigma_{a_2}^2 \end{pmatrix}, \quad (10)$$

where $\sigma_{a_1}^2$ and $\sigma_{a_2}^2$ are the additive genetic variances for traits 1 and 2, and σ_{a_1, a_2} is their additive genetic covariance. The joint phenotypic covariance is $\text{Cov}(\mathbf{y}) = \mathbf{Z}(\mathbf{G}_0 \otimes \mathbf{\Sigma})\mathbf{Z}^T + \mathbf{R}$, where $\mathbf{R}$ is the residual covariance matrix defined analogously to $\mathbf{G}_0$ but with error (co)variances. For the categorical trait (trait 2), a threshold model is assumed on the underlying liability scale ([Sorensen and Gianola 2002](#)), where the observed categorical response is determined by an underlying continuous liability exceeding threshold values. Prior distributions were assigned following the default structure of JWAS ([Cheng et al. 2018](#)). Fixed effects intercept, trial (site), and plot nested within trial received flat, improper priors $p(\beta) \propto 1$, placing no constraint on location parameters. The plot-within-trial term was treated as an i.i.d. random effect with a scaled inverse chi-squared prior on its variance, with degrees of freedom $\nu = 4.0$ (JWAS default) and scale derived from the phenotypic variance of the corresponding trait.

For the multivariate additive genetic component under GBLUP, the genetic covariance matrix $\mathbf{G}_0$ was assigned an inverse-Wishart prior,

$$\mathbf{G}_0 \sim \mathcal{IW}(\mathbf{V}_0, \nu_0), \quad (11)$$

with degrees of freedom $\nu_0 = 4.0$ and scale matrix $\mathbf{V}_0$ derived from the phenotypic (co)variances of the response traits, consistent with the JWAS default when no explicit prior scale is supplied ([Cheng et al. 2018](#)). The residual covariance matrix $\mathbf{R}$ was assigned an analogous inverse-Wishart prior with the same default parameterisation. No informative prior knowledge was encoded for any variance component.

For the H-matrix model, threshold formulation was assumed for the categorical trait Lev17 (tree survival, binary) on the underlying liability scale ([Sorensen and Gianola 2002](#)). Observed binary responses were recoded to category labels {1, 2} as required by JWAS, and liability values were imputed at each Gibbs iteration by data augmentation ([Cheng et al. 2018](#)). The liability-scale residual variance for Lev17 was fixed at $\sigma_e^2 = 1$ for model identifiability, which is the standard constraint in threshold models and removes one degree of freedom from the residual variance prior for that trait; the threshold separating the two liability categories was fixed at zero. The three continuous traits (Hjd17, Htv17, Sprant17) retained freely estimated residual variances under the same inverse-Wishart prior structure as the G-matrix model.

After inferring breeding values $\mathbf{a}_1$ and $\mathbf{a}_2$, a joint selection index matrix $\mathbf{I}$ of size $n \times l$ was defined for all members of the n -individual pedigree across l MCMC iterations by weighting the breeding values of the traits. For simplicity, all included traits were given equal weights. The multi-trait mixed effects model was implemented in the JWAS package ([Cheng et al. 2018](#)) in Julia using default settings, with the total chain length of 52,000 iterations, a burn-in of 2,000 iterations, and thinning by a factor of 10, yielding $l = 5,000$ post-burn-in posterior samples for all subsequent OCS analyses. Chain convergence was assessed by visual inspection of trace plots and by computing the effective sample size (ESS) for all heritability and genetic variance parameters using the MCMCDiagnosticTools.jl package; convergence statistics are reported in [Supplementary Table S1](#).

MCMC-based uncertainty analysis

Before introducing the CVaR-OCS optimization framework, we first characterize how EBV uncertainty propagates into

contribution decisions by solving OCS separately for each of the l MCMC posterior draws. This repeated sampling procedure (referred to as MCMC-OCS throughout) serves as a diagnostic tool. The resulting distribution of contribution vectors $\{\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(l)}\}$ reveals which MAP-OCS selections are stable across the posterior and which are sensitive to EBV uncertainty, directly motivating the single-solve CVaR-OCS formulation described in Section Conditional Value at Risk. Formally, for each MCMC iteration j we solve:

$$\mathbf{c}^{(j)} = f(\mathbf{g}^{(j)} = \frac{1}{2}(\mathbf{a}_1^{(j)} + \mathbf{a}_2^{(j)}), \boldsymbol{\Sigma}), \quad j = 1, \dots, l, \quad (12)$$

where $f(\cdot)$ is the OCS function call and $\mathbf{a}_1^{(j)}, \mathbf{a}_2^{(j)}$ are the EBVs of the selection candidates for the continuous and categorical traits in iteration j . For the general m -trait case the index weights $\mathbf{w}$ replace the $\frac{1}{2}$ scalar. Using this procedure, we acquire a matrix of contributions $\mathbf{C} = [\mathbf{c}^{(1)} \mathbf{c}^{(2)} \dots \mathbf{c}^{(l)}]$.

MAP-OCS reference solution

The reference optimum contribution vector $\mathbf{c}^{\text{MAP}}$ was computed using maximum a posteriori (MAP) breeding values:

$$\mathbf{c}^{\text{MAP}} = \operatorname{argmax}_{\mathbf{c}} \left\{ (\mathbf{g}^{\text{MAP}})^T \mathbf{c} \right\} \quad (13)$$

subject to constraints (2)–(7), where $\mathbf{g}^{\text{MAP}}$ is the vector of MAP breeding values.

Overlap and agreement metrics

For each MCMC iteration j , the top 100 contributors were identified and the overlap with the MAP-OCS solution was computed as

$$\text{Overlap}_j = |\mathcal{S}^{\text{MAP}} \cap \mathcal{S}^{(j)}|, \quad (14)$$

where $\mathcal{S}^{\text{MAP}}$ denotes the set of top 100 MAP-OCS contributors and $\mathcal{S}^{(j)}$ denotes the top 100 contributors from iteration j . Jaccard similarity was computed as

$$J_j = \frac{|\mathcal{S}^{\text{MAP}} \cap \mathcal{S}^{(j)}|}{|\mathcal{S}^{\text{MAP}} \cup \mathcal{S}^{(j)}|}. \quad (15)$$

EBV-selection frequency analysis

Selection frequency for each individual across all MCMC iterations was calculated as

$$f_i = \frac{1}{l} \sum_{j=1}^l \mathbf{1}(i \in \mathcal{S}^{(j)}), \quad (16)$$

where $\mathbf{1}(i \in \mathcal{S}^{(j)})$ is an indicator function equal to 1 if individual i is in the top 100 contributors in iteration j . The relationship between MAP EBV and MCMC-OCS selection frequency was analysed using Gaussian process (GP) regression (Rasmussen and Williams 2006) with a Matérn 3/2 covariance function, with all hyperparameters estimated via maximum likelihood using the Nelder–Mead algorithm. The GP model was fitted using AbstractGPs.jl (van der Wilk et al. 2020) and Optim.jl (Mogensen and Riseth 2018) in Julia 1.11.

Stochastic programming interpretation of MCMC-OCS

The OCS problem (1)–(7) treats the breeding value vector $\mathbf{g}$ as a fixed, known quantity. In reality $\mathbf{g}$ is a random vector whose

uncertainty is fully characterised by the posterior distribution $p(\mathbf{g} | \mathbf{y})$ obtained from the Bayesian MCMC analysis described above. Ignoring this uncertainty can lead to suboptimal and fragile selection decisions (Pocrnic et al. 2023; Fogg et al. 2024). A natural framework for making decisions under such uncertainty is *stochastic programming* (Shapiro et al. 2009; Birge and Louveaux 2011), in which uncertain parameters are represented by a probability distribution and the optimisation objective is defined as an expectation (or a risk-adjusted expectation) over that distribution.

Scenario-based formulation

Let Ω denote the probability space of breeding values and let $\boldsymbol{\xi}(\omega) = \mathbf{g}(\omega)$ be the random breeding-value vector for outcome $\omega \in \Omega$. The general mean-risk stochastic OCS problem is

$$\text{maximize}_{\mathbf{c}} \quad \mathbb{E}_{\omega}[\mathbf{c}^T \boldsymbol{\xi}(\omega)] - \lambda \frac{1}{2} \mathbf{c}^T \boldsymbol{\Sigma} \mathbf{c} - \mu \rho(\mathbf{c}^T \boldsymbol{\xi}(\omega)), \quad (17)$$

subject to the feasibility constraints (3)–(7), where $\lambda > 0$ is the inbreeding-penalty weight replacing the hard constraint (2) with a Lagrangian penalty (these two formulations trace the same efficient frontier as λ varies), $\mu \geq 0$ is a risk-aversion weight, and $\rho(\cdot)$ is a chosen risk measure applied to the distribution of realised genetic gain $\mathbf{c}^T \boldsymbol{\xi}(\omega)$.

The MCMC sampler produces l approximately independent draws $\{\mathbf{g}^{(1)}, \mathbf{g}^{(2)}, \dots, \mathbf{g}^{(l)}\}$ from the posterior $p(\mathbf{g} | \mathbf{y})$. Each draw constitutes a scenario, and the Sample Average Approximation (SAA) (Shapiro et al. 2009; Klein Haneveld et al. 2015) replaces the expectation and the risk measure by their empirical counterparts over the l scenarios

$$\text{maximize}_{\mathbf{c}} \quad \frac{1}{l} \sum_{j=1}^l \mathbf{c}^T \mathbf{g}^{(j)} - \lambda \frac{1}{2} \mathbf{c}^T \boldsymbol{\Sigma} \mathbf{c} - \mu \hat{\rho} \left(\left\{ \mathbf{c}^T \mathbf{g}^{(j)} \right\}_{j=1}^l \right), \quad (18)$$

subject to (3)–(7), where $\hat{\rho}$ is the empirical risk measure over the l scenarios. When $\mu = 0$ the objective reduces to $\mathbf{c}^T \bar{\mathbf{g}}$, where $\bar{\mathbf{g}} = l^{-1} \sum_{j=1}^l \mathbf{g}^{(j)}$ is the posterior mean EBV, and the problem is reduced to a standard QP equivalent to MAP-OCS evaluated at the posterior mean.

Relationship to the MCMC-OCS procedure. The MCMC-OCS procedure implemented in the present study (following Pocrnic et al. 2023) solves l separate OCS problems, one per MCMC draw (12), and analyses the resulting distribution of contribution vectors $\{\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(l)}\}$. This differs fundamentally from the joint SAA problem (18). In our procedure the contribution vector $\mathbf{c}$ is re-optimised for each scenario rather than held fixed across all scenarios, so EBV uncertainty appears in the output distribution of $\mathbf{c}$ but not in the optimisation itself. The joint SAA (18), described here for theoretical completeness, identifies a single robust $\mathbf{c}$ that explicitly trades off expected gain, inbreeding and estimation risk simultaneously. This formulation is the basis for the CVaR-OCS extension introduced below.

Connection to Markowitz mean–variance theory

Formulation (18) has an instructive connection to Markowitz mean–variance portfolio theory (Markowitz 1952, 1959). In the Markowitz problem, the variance constraint is relaxed into the objective via its Lagrange multiplier, yielding the penalised form

$$\text{maximize}_{\mathbf{c}} \quad \mathbf{c}^T \boldsymbol{\mu}_r - \lambda \mathbf{c}^T \boldsymbol{\Sigma}_r \mathbf{c}, \quad (19)$$

where $\boldsymbol{\mu}_r$ is the vector of expected returns and $\boldsymbol{\Sigma}_r$ is their covariance matrix. The OCS objective (18) (with $\mu = 0$) is structurally identical, with the posterior mean $\mathbf{g}$ playing the role of expected returns and $\boldsymbol{\Sigma}$ playing the role of the return covariance matrix. The dual formulation, which maximises the expected genetic gain for a given inbreeding budget, traces the same efficient frontier. In addition, the Lagrange multiplier λ is the risk of inbreeding, analogous to the market price of risk in finance. This connection motivates the portfolio-level vulnerability metrics described in the [Supplementary Material](#), which are conceptually equivalent to portfolio diversification measures ([Markowitz 1952](#)).

Value at risk

When breeding values are uncertain, the genetic gain $Z(\mathbf{c}, \boldsymbol{\xi}) = \mathbf{c}^T \boldsymbol{\xi}$ is itself a random variable. A natural summary of the downside of this distribution is the *Value at Risk* (VaR) at confidence level $\alpha \in (0, 1)$, defined as the α -quantile of the gain distribution

$$\text{VaR}_\alpha(\mathbf{c}) = \inf\{\eta \in \mathbb{R} : \text{Pr}[Z(\mathbf{c}, \boldsymbol{\xi}) < \eta] \leq 1 - \alpha\}. \quad (20)$$

Intuitively, $\text{VaR}_\alpha(\mathbf{c})$ is the level of genetic gain that will be exceeded with probability α . For example, with $\alpha = 0.95$, the VaR is the genetic gain that will not be achieved only 5% of the time across the posterior distribution of breeding values. In the OCS context, a low VaR_α for a given contribution vector $\mathbf{c}$ signals that the selection decision is sensitive to EBV uncertainty. Even moderate perturbations of the breeding values can substantially reduce genetic gain. In the SAA setting, VaR_α is estimated by the empirical α -quantile of the sample $\{\mathbf{c}^T \mathbf{g}^{(j)}\}_{j=1}^l$.

Despite its intuitive appeal, VaR has well-known limitations as a risk measure. It conveys no information about the severity of outcomes beyond the threshold. Furthermore, VaR is not in general a coherent risk measure ([Artzner et al. 1999](#)) since it lacks the subadditivity property, so combining two contribution vectors can increase rather than decrease VaR, which is an undesirable feature for a diversification-based framework such as OCS.

Conditional value at risk

These limitations are addressed by the *Conditional Value at Risk* (CVaR), also known as Expected Shortfall ([Rockafellar and Uryasev 2000, 2002](#)). CVaR at level α is defined as the expected genetic gain conditional on being in the worst $(1 - \alpha)$ fraction of the posterior distribution:

$$\text{CVaR}_\alpha(\mathbf{c}) = \mathbb{E}[Z(\mathbf{c}, \boldsymbol{\xi}) | Z(\mathbf{c}, \boldsymbol{\xi}) \leq \text{VaR}_\alpha(\mathbf{c})]. \quad (21)$$

CVaR is a coherent risk measure ([Artzner et al. 1999](#)) and is therefore well suited to the OCS diversification objective. Maximising $\text{CVaR}_\alpha(\mathbf{c})$ is equivalent to maximising the expected genetic gain in the worst $(1 - \alpha)$ scenarios, providing explicit protection against cases where the true breeding values are substantially lower than their posterior means.

A key computational advantage of CVaR is the [Rockafellar and Uryasev \(2000\)](#) auxiliary-variable representation

$$\text{CVaR}_\alpha(\mathbf{c}) = \max_{\eta \in \mathbb{R}} \left\{ \eta - \frac{1}{1 - \alpha} \mathbb{E}[\max(0, \eta - Z(\mathbf{c}, \boldsymbol{\xi}))] \right\}, \quad (22)$$

where η equals $\text{VaR}_\alpha(\mathbf{c})$ at the optimum. Substituting the SAA estimator for the expectation and reinstating the hard inbreeding constraint gives the following tractable QP for CVaR-aware OCS

$$\text{maximize}_{\mathbf{c}, \boldsymbol{\eta}, \mathbf{z}} \quad \frac{1}{l} \sum_{j=1}^l \mathbf{c}^T \mathbf{g}^{(j)} + \mu \left[\eta - \frac{1}{(1 - \alpha)l} \sum_{j=1}^l z_j \right], \quad (23)$$

$$\text{subject to} \quad z_j \geq \eta - \mathbf{c}^T \mathbf{g}^{(j)}, \quad j = 1, \dots, l, \quad (24)$$

$$z_j \geq 0, \quad j = 1, \dots, l, \quad (25)$$

$$\frac{1}{2} \mathbf{c}^T \boldsymbol{\Sigma} \mathbf{c} \leq \boldsymbol{\Theta}, \quad (26)$$

$$\mathbf{c}^T \mathbf{s} = 0.5, \quad \mathbf{c}^T \mathbf{d} = 0.5, \quad (27)$$

$$\mathbf{c}^T \mathbf{1} = 1, \quad (28)$$

$$\mathbf{0} \leq \mathbf{c} \leq \mathbf{m}. \quad (29)$$

Here $z_j = \max(0, \eta - \mathbf{c}^T \mathbf{g}^{(j)})$ is the shortfall of genetic gain in scenario j below the VaR threshold η , and $\mu \geq 0$ controls the weight given to tail-risk protection. Note that the inbreeding penalty of (17) has been replaced by the hard constraint (26) with threshold $\boldsymbol{\Theta}$, consistent with the original OCS formulation (2). Thus μ is therefore the sole risk-aversion parameter. As in (3)–(4), the gender constraints (27) are included only for dioecious species.

Problem (23)–(29) is a QP in $(\mathbf{c}, \boldsymbol{\eta}, \mathbf{z})$, quadratic in $\mathbf{c}$ through (26), and linear in z_j and η , and can be solved directly with the COSMO solver ([Garstka et al. 2021](#)) in JuMP ([Dunning et al. 2017](#)). The objective has a clear two-component interpretation: the first term maximises expected genetic gain across all posterior scenarios (analogous to the expected-return term in Markowitz), while the CVaR term penalises poor outcomes in the tail of the breeding-value distribution, capturing candidate-specific estimation risk. As $\mu \rightarrow 0$ the CVaR term vanishes and the problem reduces to MAP-OCS at the posterior mean, while as $\alpha \rightarrow 1$, CVaR converges to the worst-case scenario, approaching the minimax robust OCS of [Fogg et al. \(2024\)](#). The parameter μ therefore spans a risk–return continuum that generalises both MAP-OCS and the robust OCS of [Fogg et al. \(2024\)](#) as limiting cases.

CVaR-OCS parameter selection

The CVaR-OCS formulation (23)–(29) involves two parameters that jointly determine the operating point on the risk–return efficiency frontier, the confidence level $\alpha \in (0, 1)$ and the risk-aversion weight $\mu \geq 0$.

The confidence level α defines which tail of the gain distribution is protected. At $\alpha = 0.90$, CVaR-OCS optimises the expected genetic gain in the worst 10% of MCMC scenarios, and at $\alpha = 0.99$ it focuses exclusively on the worst 1%. Higher α therefore provides stronger tail protection but concentrates the optimisation objective on an increasingly small and extreme subset of posterior draws, which can reduce stability when the number of MCMC iterations l is modest. The risk-aversion weight μ governs the relative importance of tail-risk protection versus expected genetic gain. At $\mu = 0$ the CVaR term vanishes and the problem reduces to MAP-OCS evaluated at the posterior mean (equation 23), while increasing μ progressively shifts the solution toward the minimax worst-case limit ([Fogg et al. 2024](#)).

To select a practically useful operating point, we evaluated all combinations of $\alpha \in \{0.90, 0.95, 0.99\}$ and $\mu \in \{0.50, 0.80, 1.00, 1.25, 1.50, 1.75, 2.00, 5.00, 10.00\}$ for each dataset. For each (α, μ) combination, we recorded the percentage genetic gain loss and the percentage CVaR $_{1-\alpha}$ improvement relative to

Table 1. Description of included datasets. Average coancestry (Avg. coanc.) of the candidate population is calculated via the pedigree.

Study	Species	Pedigree	Candidates	Gen.	Parents	Families	Avg. coanc.
Szydlowski and Paczyńska (2011)	simulated animals	3,226	900	5	59	30	0.054
Chen et al. (2018)	<i>P. abies</i>	5,597	5,525	2	72	142	0.016
Chen et al. (2018)	<i>P. abies</i>	1,273	1,218	2	55	126	0.014
Resende et al. (2012)	<i>P. taeda</i>	973	926	2	43	71	0.043

the MAP-OCS baseline, yielding a gain–tail-risk efficiency frontier for each α level.

The recommended operating point on each frontier was identified using a geometric perpendicular-distance criterion (see also Satopaa et al. 2011). Each frontier was normalised to [0, 1] in both dimensions and the point of maximum perpendicular distance from the line connecting the two endpoints was taken as the elbow, i.e. the point beyond which additional tail-gain improvement comes at disproportionately high genetic gain cost. As a confirmation, we computed the marginal efficiency ratio between successive μ values, defined as the ratio of $CVaR_{1-\alpha}$ improvement to loss of genetic gain. A sharp drop in this ratio confirms that the elbow marks a genuine change in the efficiency regime rather than a gradual transition. The selected (α, μ) combination was used for all downstream comparisons within each dataset.

MCMC-based robustness scores and contribution distribution

To characterize which MAP-OCS selections are most vulnerable to EBV uncertainty, we calculated individual robustness scores by quantifying the expected loss of genetic gain when each candidate is excluded from the selected cohort across MCMC iterations. We also expressed gain as a percentage relative to the mean baseline, providing a more stable normalized metric. High-risk individuals were then identified using a dynamic bottom-quartile threshold applied to MAP-OCS selections. To characterize population-level genetic gain concentration, we computed four complementary contribution distribution metrics: mean marginal gain, maximum marginal gain, top-10% concentration ratio, and the Gini coefficient. Full definitions, equations, and computational details are provided in the [Supplementary Material](#).

Simulated animal breeding dataset

The simulated dataset from the 14th QTL-MAS workshop (Szydlowski and Paczyńska 2011) was used to validate CVaR-OCS in an animal breeding context with explicit sex structure and known true breeding values (TBVs). The pedigree comprised $n = 3,226$ individuals across five generations, founded by 5 sires and 15 dams. Two correlated traits, a quantitative trait and a binary trait, were simulated under a model including additive, epistatic, and parent-of-origin effects.

Phenotypic records were available for $n_{\text{pheno}} = 2,326$ individuals from the first four generations, while the $n_{\text{cand}} = 900$ individuals of the fifth generation were un-phenotyped and served as the candidate population for OCS, mimicking a standard genomic selection scenario in which selection candidates lack own performance records. Genotypic data comprised $p = 9,723$ SNPs on five chromosomes after quality control (MAF > 0.05 , missingness $< 10\%$). Sex was recorded for all individuals and included as a fixed effect in the Bayesian model.

GBVs and their full posterior distributions were estimated for all $n = 3,226$ individuals using a bivariate Bayesian model in JWAS (Cheng et al. 2018), with the genomic relationship matrix $\mathbf{G}$

constructed following VanRaden (2008) for use in the OCS coancestry constraint. The posterior MCMC samples of GBVs for the $n_{\text{cand}} = 900$ fifth-generation candidates were subsequently used as input to both MAP-OCS (using the posterior mean as a point estimate) and MCMC-OCS. Because the true additive breeding values of all individuals are known by construction in this dataset, the selections recommended by each method could be evaluated against the TBV-optimal solution, which provides a direct assessment of decision quality under EBV uncertainty.

Norway spruce case study

Two field trials of Norway spruce, with Skogforsk IDs S23F8820476 located outside Vindeln, Västerbotten County (64.3°Latitude, 19.68°Longitude) and S23F8820477, Hälsanberg, Västernorrland County (63.58°Latitude, 18.20°Longitude), were utilized for illustrating CVaR-OCS and MCMC-OCS on forest tree breeding field trials (Table 1). In both trials, approximately 6,000 offspring from 150 families were planted in 1988 in single tree plot design for evaluating the performance of parent trees in the Hissjö seed orchard. The following phenotypic traits were scored: total tree height at year 7 and 17, early height growth at year 17, number of spike knots at year 17 and tree vitality at year 7. Of those traits, tree height, height growth and number of ramcorns at age 17 were used in the selection index with equal weight (negative weight for ramcorn number). In total, 1,218 offspring (568 and 650 in S23F8820476 and S23F8820477 respectively) were genotyped using exome capture technique resulting in 300 714 SNPs prior to filtration. For further details of the Norway spruce data, please consult Chen et al. (2018); a summary of the phenotypic data is provided in [Supplementary Table S2](#). The $\mathbf{G}$ matrix was calculated using the rrBLUP package (Endelman 2011), with default settings. To combine genotyped and ungenotyped trees, a joint $\mathbf{H}$ matrix was calculated using the AGHmatrix package (Amadeu et al. 2023), where $\mathbf{G}$ was first tuned towards $\mathbf{A}$ following Christensen et al. (2012) by matching mean diagonal and off-diagonal elements (tuning coefficients $\alpha = 1.025$, $\beta = 0.029$). The blending was performed using the Hmatrix function, with $\tau = 1$ and $\omega = 1$ (default) (Legarra et al. 2009). To ensure that $\mathbf{H}$ used in subsequent analyses was positive definite, we applied the cor_nearPD function from the Bigsimr.jl package. Multi-trait analysis to obtain posterior estimates of breeding values (BV) were performed using the JWAS Julia package (Cheng et al. 2018) with a total chain length of 52,000 iterations, a burn-in of 2,000 iterations, and thinning by a factor of 10, yielding $l = 5,000$ post-burn-in posterior samples. The predictors used in the model were a site factor and a nested plot within site term.

Loblolly pine case study

As an additional empirical case study, CVaR-OCS was also applied to a Loblolly pine (*Pinus taeda*) breeding population (Lower Gulf Elite Population) in the southeastern United States (Resende et al. 2012; Isik and McKeand 2019) ($n = 926$ genotyped candidates; Table 1), using the same posterior-sampling and optimization pipeline described above. Full dataset description, analysis, and

results for this case study are provided in the [Supplementary Material](#), including MCMC convergence diagnostics ([Supplementary Figs. S3–S4](#)) and complete CVaR-OCS versus MAP-OCS results ([Supplementary Table S10](#)).

Results

Validation on a simulated dataset with known true breeding values

To evaluate the new proposed framework CVaR-OCS under controlled conditions where true breeding values (TBVs) are known, we analysed the QTL-MAS 2010 workshop dataset ([Szydlowski and Paczyńska 2011](#)), which provides a realistic multi-generation genomic selection scenario with a fully genotyped pedigree. The final generation of $n = 900$ unphenotyped selection candidates (433 sires, 467 dams) served as the target population for OCS, with a primary coancestry constraint of $\Theta = 0.03$. Posterior distributions were obtained from a two-trait Bayesian GBLUP model fitted in JWAS with $l = 5,000$ post-burnin MCMC samples ([Supplementary Figs. S1–S2](#)). An oracle solution, MAP-OCS(TBV), computed using known TBVs in place of EBVs, served as the theoretical optimum against which both EBV-based methods were evaluated. The bivariate JWAS model estimated moderate heritabilities for both traits ($\hat{h}_{y_1}^2 = 0.46 \pm 0.017$ for the continuous trait; $\hat{h}_{y_2}^2 = 0.47 \pm 0.030$ for the binary trait, fitted as a threshold model) and a positive inferred genetic correlation of $\hat{r}_g = 0.69$, indicating that a combined selection index with equal weights captures genetic merit across both simultaneously. These estimates closely recovered the true simulated values reported for this dataset ($h^2 = 0.39$ – 0.52 for the quantitative trait, sex-specific, and 0.48 for the binary trait; genetic correlation 0.59 – 0.68 ; [Szydlowski and Paczyńska 2011](#)), confirming that the model correctly recovered the underlying genetic architecture.

MCMC-OCS characterizes the impact of EBV uncertainty on contributions

Before evaluating CVaR-OCS, we first characterised how EBV uncertainty propagates into contribution decisions using MCMC-OCS: MAP-OCS was independently re-solved against each of the $l = 5,000$ posterior EBV scenarios (sex-constrained, $\Theta = 0.03$), and the resulting per-scenario solutions were compared to the MAP-OCS(EBV) reference solution. Across scenarios, mean overlap with the MAP-OCS(EBV) selected set was 13.4 individuals (range: 5–24; MAP-OCS(EBV) itself selected 29), with a mean Jaccard similarity of 27.4%, indicating that a majority of per-scenario optimal solutions differ substantially from the single point-estimate solution. Selection frequency across the $l = 5,000$ solutions was strongly rank-correlated with the MAP EBV index (Spearman $\rho = 0.875$): candidates in the top EBV quartile were selected in an average of 13.0% of scenarios (maximum 72.2%), whereas 229 of the 900 candidates (25.4%), concentrated in the bottom EBV quartile, were never selected in any of the 5,000 independent re-optimisations, which gives a top-to-bottom selection-frequency ratio of 829 $\times$. This heterogeneity in per-scenario selection reflects genuine uncertainty in which candidates are truly optimal given the posterior distribution of breeding values, motivating an approach such as CVaR-OCS that accounts for this uncertainty directly within a single optimisation rather than requiring repeated re-solving.

CVaR-OCS parameterisation and elbow selection

To identify the operating point that appropriately penalises EBV uncertainty without sacrificing excessive genetic gain, we

evaluated all combinations of $\alpha \in \{0.90, 0.95, 0.99\}$ and $\mu \in \{0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 5.0, 10.0\}$ across three coancestry constraints $\Theta \in \{0.02, 0.03, 0.05\}$, with results at $\Theta = 0.03$ presented in the main text and the remaining constraints in [Supplementary Table S7](#). The efficiency frontier ([Fig. 1a](#)), expressed as percentage genetic gain loss and CVaR₉₅ improvement relative to the MAP-OCS in-sample baseline, reveals a consistent concave relationship. Early increments in μ yield disproportionate tail-gain improvements, with returns diminishing beyond a species-specific elbow.

Elbow points were identified using the geometric perpendicular-distance criterion (CVaR-OCS *parameter selection* in Materials and Methods): $\alpha = 0.90$ at $\mu = 2.00$ (CVaR₉₅ +0.46%, gain loss +0.07%), $\alpha = 0.95$ at $\mu = 1.25$ (CVaR₉₅ +0.69%, gain loss –0.12%), and $\alpha = 0.99$ at $\mu = 2.00$ (CVaR₉₅ +0.62%, gain loss +0.18%; [Table 2](#)). The recommended operating point is $\alpha = 0.95$, $\mu = 1.25$, which achieves the highest CVaR₉₅ improvement (+0.69%) at negligible gain cost (–0.12% vs. the MAP-OCS in-sample baseline; [Fig. 1a](#)). Across the three confidence levels, the elbow-point CVaR₉₅ improvement was highest at $\alpha = 0.95$; $\alpha = 0.99$'s tail estimate is based on only the worst 1% of the $l = 5,000$ scenarios (50 draws), which likely contributes to its lower improvement relative to $\alpha = 0.95$ despite exceeding $\alpha = 0.90$.

Tail-risk protection and contribution structure

Having selected $\alpha = 0.95$, $\mu = 1.25$ as the recommended operating point, we examined what CVaR-OCS actually delivers relative to MAP-OCS in terms of tail-gain security and the structure of parental contributions. At $\alpha = 0.95$, $\mu = 1.25$, CVaR-OCS improved CVaR₉₅ by 0.69% relative to MAP-OCS(EBV) ($1.400 \rightarrow 1.409$) and VaR₉₅ by 0.58% ($1.440 \rightarrow 1.448$) when both solutions were evaluated across all $l = 5,000$ MCMC scenarios ([Fig. 1d](#); [Table 2](#)). This tail-gain improvement was obtained with a marginal change in expected genetic gain ($1.595 \rightarrow 1.596$), indicating that CVaR-OCS compresses the lower tail of the gain distribution with little effect on its centre. CVaR-OCS selected 35 parents compared to 29 under MAP-OCS (+20.7% broader selection base), distributing contributions somewhat more equitably (Gini: 0.406 vs. 0.347). All 29 MAP-OCS selections were retained by CVaR-OCS, which additionally recruited 6 new individuals ([Fig. 1b](#)); no MAP-OCS selection was dropped.

Estimation bias in MAP-OCS

MAP-OCS optimises contributions against posterior-mean EBVs, treating them as fixed, known quantities. Because these posterior means are imperfect predictors of the true breeding values (index accuracy $r = 0.76$ against TBVs in this dataset), the contribution vector that maximises gain under the EBVs does not maximise realised gain on the true values. At $\Theta = 0.03$, MAP-OCS(EBV) realised a true-breeding-value gain of 1.34, only 78% of the 1.72 achieved by the oracle solution that optimises directly against true values (*Oracle comparison*, below). This shortfall quantifies the cost of treating uncertain EBVs as if they were known. CVaR-OCS attains an in-sample expected gain (1.596 at $\alpha = 0.95$, $\mu = 1.25$) essentially identical to that of MAP-OCS (1.595); its tail-risk protection is therefore obtained at negligible expected-gain cost in-sample, though its realised gain against true breeding values (1.332, a 0.73% reduction relative to MAP-OCS).

Oracle comparison

The oracle MAP-OCS(TBV) selected 31 individuals spanning a wider EBV range than either EBV-based method ([Fig. 1b,c](#)), reflecting the oracle's knowledge that several high-EBV individuals carry

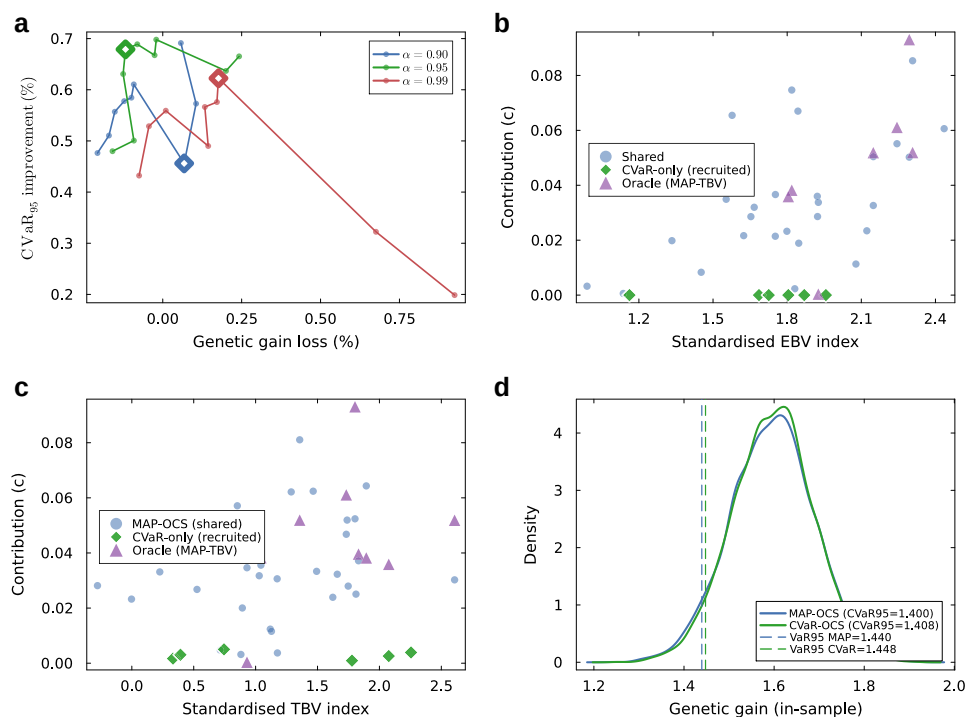


Fig. 1. CVaR-OCS validation on the QTLMAS 2010 simulated dataset ($n = 900$ candidates; $\Theta = 0.03$). a) Gain-tail-risk efficiency frontier across three confidence levels ($\alpha \in \{0.90, 0.95, 0.99\}$) and CVaR weights ($\mu \in \{0.50, \dots, 10.00\}$), expressed as percentage genetic gain loss and CVaR₉₅ improvement relative to the MAP-OCS in-sample baseline. Open diamonds mark the elbow point for each α level, identified by the geometric perpendicular-distance criterion (CVaR-OCS parameter selection in Materials and Methods), beyond which additional CVaR₉₅ improvement comes at disproportionate genetic gain cost. b) Standardised EBV index vs parental contribution for MAP-OCS and the recommended CVaR-OCS solution ($\alpha = 0.95$, $\mu = 1.25$), coloured by selection group: individuals shared between MAP-OCS and CVaR-OCS (blue circles), and individuals newly recruited by CVaR-OCS (green diamonds). The oracle solution based on true breeding values (MAP-OCS(TBV); purple triangles) is shown for reference. c) Standardised TBV index vs parental contribution for the same three solutions as in panel b). CVaR-OCS recruits individuals whose true breeding values more closely align with the oracle solution compared to MAP-OCS (EBV-based), demonstrating improved robustness to EBV uncertainty when true genetic merit is considered. d) Kernel density of genetic gain realised by MAP-OCS (blue) and CVaR-OCS (green) evaluated across all $l = 5,000$ MCMC EBV scenarios. Dashed vertical lines mark the 5th-percentile gain (VaR₉₅) for each solution; CVaR₉₅ values are given in the legend.

Table 2. CVaR-OCS results for the QTL-MAS 2010 simulation study at the recommended operating point ($\Theta = 0.03$; $\alpha^* = 0.95$, $\mu^* = 1.25$; $n = 900$ candidates; $l = 5,000$ MCMC scenarios). E[gain]: in-sample expected genetic gain across MCMC scenarios, on a common scale for all EBV-based solutions^a; realised gain: cTt evaluated against true breeding values (TBVs); CVaR₉₅: expected gain in the worst 5% of MCMC scenarios; n_{sel} : number of selected parents (sires S, dams D); jaccard: set overlap with oracle. Results across all three coancestry constraints ($\Theta \in \{0.02, 0.03, 0.05\}$) are provided in Supplementary Table S7.

Metric	Solution	Value	$\Delta(\%)$
E[gain]	MAP-OCS(EBV)	1.595	—
	CVaR-OCS	1.596	+0.08
	Oracle	1.725	+28.6
Realised gain	MAP-OCS(EBV)	1.342	—
	CVaR-OCS	1.332	-0.73
	Oracle	1.725	+28.6
CVaR ₉₅	MAP-OCS(EBV)	1.400	—
	CVaR-OCS	1.409	+0.69
	Oracle	1.725	+28.6
VaR ₉₅	MAP-OCS(EBV)	1.440	—
	CVaR-OCS	1.448	+0.6
n_{sel} (S, D)	Oracle	31 (16, 15)	—
	MAP-OCS(EBV)	29 (12, 17)	—
	CVaR-OCS	35 (15, 20)	+20.7
Jaccard vs oracle	MAP-OCS(EBV)	0.176	—
	CVaR-OCS	0.179	+1.7
	Oracle	0.176	—
Shared with oracle	MAP-OCS(EBV)	9	—
	CVaR-OCS	10	+11.1
	Oracle	9	—

^a E[gain] is the expected genetic gain across the l MCMC scenarios, computed on a common (per-draw standardised) scale for the EBV-based solutions so that MAP-OCS and CVaR-OCS are directly comparable. The realised gain against true breeding values (Realised gain rows) shows that MAP-OCS(EBV) recovers 78% of the oracle-achievable gain at $\Theta = 0.03$.

inflated posterior-mean estimates. Jaccard similarity between MAP-OCS(EBV) and the oracle was 0.176 ($n = 9$ shared individuals) at $\Theta = 0.03$. CVaR-OCS improved oracle overlap marginally to 0.179 ($n = 10$ shared; +11% more oracle-optimal individuals recovered), despite having no access to true values.

Application to an operational Norway spruce breeding population

Genetic parameter estimates

Heritability estimates varied markedly across relationship matrices and traits in Norway spruce (Supplementary Table S3). MCMC convergence was confirmed for all models, with chains sampling their respective stationary distributions as shown in the trace plots and posterior distributions (Supplementary Figs. S5–S6 for the G-matrix model and Supplementary Figs. S7–S8 for the H-matrix model). Genomic (G-matrix) heritabilities for the two primary height traits entering the selection index were nearly half the corresponding blended (H-matrix) estimates: $\hat{h}_G^2 = 0.251 \pm 0.048$ vs. $\hat{h}_H^2 = 0.442 \pm 0.045$ for Htv17, and $\hat{h}_G^2 = 0.221 \pm 0.056$ vs. $\hat{h}_H^2 = 0.442 \pm 0.040$ for Hjd17. The third index component, ramicorn number at age 17 (Sprant17), showed very low genomic heritability ($\hat{h}_G^2 = 0.094 \pm 0.040$), indicating that uncertainty is compounded across all three index traits. Genetic correlations among the growth and wood-quality traits were also estimated for both the genotyped subset and the full pedigree dataset (Supplementary Table S4; Supplementary Figs. S9–S10).

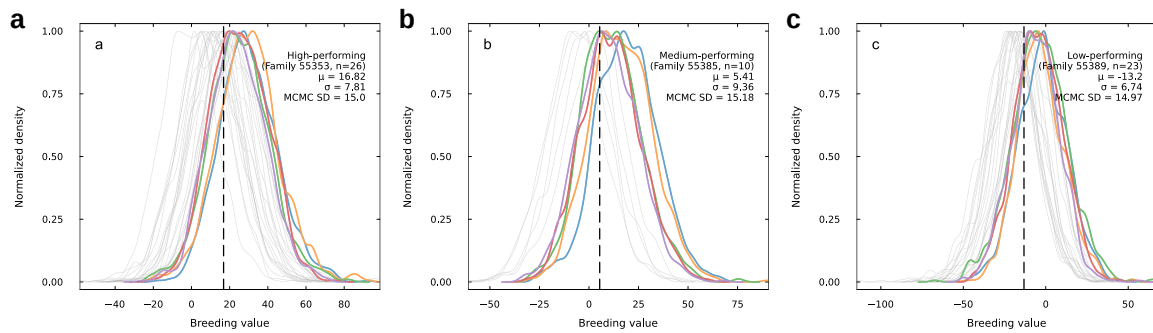


Fig. 2. Posterior uncertainty in estimated breeding values reveals ranking ambiguity within Norway spruce families. Kernel density plots show the posterior distributions of selection index breeding values (Hjd17 + Htv17 – Sprant17) for individuals within three representative families: a) high-performing (Family 55,353, $n=26$), b) medium-performing (Family 55,385, $n=10$), and c) low-performing (Family 55,389, $n=23$). Each thin curve represents the posterior distribution for one individual based on 5,000 MCMC samples. Thick curves highlight the top 5 individuals within each family ranked by posterior mean. Dashed vertical lines indicate family means.

Posterior distributions of EBVs revealed substantial ranking uncertainty within families in Norway spruce (Fig. 2). In the high-performing family, ranking standard deviation reached 7.4, with the top-ranked individual having only 48% probability of ranking in the top 10% of its family. These overlapping posterior distributions illustrate the core challenge motivating our uncertainty-aware framework.

MCMC-OCS characterizes the impact of EBV uncertainty on contributions

To characterize how EBV uncertainty propagates into OCS decisions, we solved OCS for each of the l MCMC posterior draws and compared the resulting distribution of contribution vectors to the MAP-OCS reference solution. This diagnostic procedure reveals the degree to which MAP-OCS selections are stable across the posterior distribution of breeding values and, by extension, which individuals are candidates for the risk-aware revision implemented by CVaR-OCS.

For the Norway spruce **G** population ($n=1,218$), MCMC-OCS revealed considerable uncertainty in contributions, with a mean overlap of only 26.6 individuals (range: 10–46) and mean Jaccard similarity of 15.4% relative to MAP-OCS (Supplementary Table S8; Fig. 3a–d). Nearly all individuals were selected in at least some MCMC iterations, yet not a single one was selected in more than half of them (highest at 35.7% selection frequency, Fig. 3b). Despite this instability in individual allocations, a strong rank-order relationship between EBV and selection frequency was preserved (Spearman $\rho=0.932$), with top-quartile EBV individuals selected 22.9 $\times$ more frequently than bottom-quartile individuals (Fig. 3c), confirming that MCMC-OCS correctly identifies which individuals to favour, even when precise contribution magnitudes vary across iterations. Using the **A** matrix for the same subset yielded modestly worse agreement across all metrics (Supplementary Table S8), while expanding to the full 5,525-tree population further reduced consistency (mean overlap: 14.1 [**H**] and 10.4 [**A**] individuals), reflecting greater within-family ranking ambiguity at larger sample sizes.

Robustness analysis of contributions

To motivate the CVaR-OCS approach, we first examined which MAP-OCS selections in Norway spruce are most vulnerable to EBV uncertainty using the individual robustness scores described in Materials and Methods. For the genotyped subpopulation ($n=1,218$), 40 of 145 MAP-OCS selections were identified as high-risk (bottom-quartile robustness scores), including seven of the

top-25 contributors. Hence, high contribution under MAP EBVs does not guarantee stability under posterior uncertainty. Re-running OCS after excluding these individuals improved mean robustness by 16.5% at a genetic gain cost of only 4.35%, but paradoxically increased contribution concentration (Gini +5.6%, top-10% concentration +12.9%), reflecting a structural limitation of post-hoc exclusion. Removing individuals forces the optimizer to concentrate contributions among the remaining high-EBV candidates, reintroducing fragility elsewhere in the selected population. This limitation directly motivates the suggested principled uncertainty-aware approach CVaR-OCS. Full robustness score results and risk classification figures are provided in Supplementary Fig. S11; contribution distribution metrics from the dual-metric comparison of MAP-OCS versus constrained OCS after bottom-quartile exclusion are provided in Supplementary Table S6. Robustness score computation required approximately $n_{\text{sel}} \times n_{\text{MCMC}} = 88 \times 200 = 17,600$ MAP-OCS solves for the $n=1,218$ subpopulation (ca. 1,133 min sequentially; ca. 57 min on 20 cores in parallel, see Supplementary Results), confirming that the diagnostic is practical when parallelised over candidates and/or after subsampling.

CVaR-OCS efficiency frontier and operating point selection

These results motivate a fundamentally different approach. Rather than identifying and removing uncertain individuals after optimisation, CVaR-OCS incorporates EBV uncertainty directly into the OCS objective, producing contribution vectors that are robust to the uncertainty quantified by the MCMC evaluation.

To evaluate the trade-off between expected genetic gain and resilience to EBV uncertainty, we applied CVaR-OCS to the Norway spruce genomic dataset ($n=1,218$; $\Theta=0.02$) across three confidence levels ($\alpha \in \{0.90, 0.95, 0.99\}$) and a range of CVaR weights ($\mu \in \{0.50, \dots, 10.00\}$). For each (α, μ) combination, the CVaR-OCS objective penalises solutions whose expected genetic gain deteriorates under the worst $(1-\alpha) \times 100\%$ of MCMC-sampled EBV realisations, thereby directing contributions away from individuals whose breeding values are sensitive to posterior uncertainty.

The resulting efficiency frontier (Fig. 4a) exhibits a consistent concave shape at all α levels. Small increases in μ initially produce substantial improvements in CVaR₉₅ at negligible cost to expected gain, with marginal returns declining rapidly beyond a species-specific threshold. To identify the optimal operating point for each α , we applied the geometric elbow criterion that locates the frontier point with maximum perpendicular distance from the chord connecting the MAP-OCS baseline to the maximum- μ

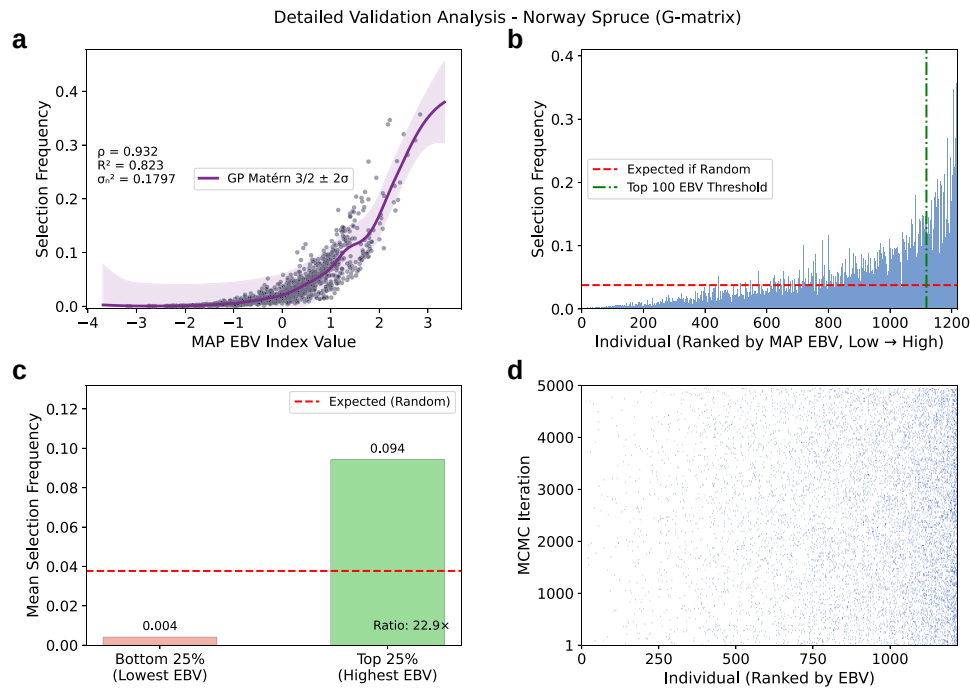


Fig. 3. Detailed validation analysis of MCMC-based optimum contribution selection uncertainty for the Norway spruce breeding population (G-matrix, $n = 1,218$). a) Relationship between MAP breeding values and selection frequency across MCMC iterations, with GP regression trend (Matérn 3/2 kernel) indicating a strong positive correlation despite high iteration-level variability. Annotation gives Spearman correlation (ρ), GP variance explained (R^2), and inferred noise variance (σ_n^2). b) Selection frequency for individuals ranked by MAP breeding value, showing a clear gradient from low-EBV (rarely selected) to high-EBV (frequently selected) individuals; red dashed line = expected random selection frequency; green dash-dot = top-100 EBV threshold. c) Quartile comparison of mean selection frequencies for the bottom vs. top 25% of breeding values; red dashed line = expected random frequency. d) Binary heatmap of selection patterns across a sample of MCMC iterations (white = not selected, blue = selected with an average 46 of individuals per iteration), with individuals sorted by breeding value rank.

solution (see CVaR-OCS *parameter selection* in Materials and Methods). Elbow points and their associated gain-robustness trade-offs are summarised in Table 3 and illustrated in Fig. 4. The sharpest and most cost-efficient elbow was observed at $\alpha = 0.90$, $\mu = 1.25$, where marginal efficiency, measured as the ratio of CVaR₉₅ improvement to genetic gain loss, dropped by 58% between $\mu = 1.25$ and $\mu = 1.50$ (from 1.84 to 0.77 $\Delta\text{CVaR}_{95}\%/\Delta\text{gain loss}\%$), confirming a pronounced change in the efficiency regime (Fig. 4, panel b). More conservative confidence levels identified their elbows at $\mu = 1.75$ ($\alpha = 0.95$ and $\alpha = 0.99$), but at gain costs of 2.06% and 3.67%, respectively: three to five times higher than the $\alpha = 0.90$ elbow. We therefore adopted $\alpha = 0.90$, $\mu = 1.25$ as the recommended operating point for all subsequent analyses.

Solution properties at the recommended operating point

At $\alpha = 0.90$, $\mu = 1.25$, CVaR-OCS achieved a CVaR₉₅ improvement of 6.60% relative to MAP-OCS (0.410 $\rightarrow$ 0.437), alongside a 5.00% improvement in VaR₉₅ (0.456 $\rightarrow$ 0.479), while reducing expected genetic gain by only 0.70% (0.644 $\rightarrow$ 0.639; Fig. 4, panel c d). This represents a highly favourable trade-off given the substantial EBV uncertainty characteristic of Norway spruce genomic evaluations, where G-matrix-based correlations are considerably weaker and more variable than pedigree-based estimates due to the relatively modest genotyped sample size.

The CVaR-OCS solution selected 159 parents compared to 145 under MAP-OCS, a 9.7% broader selection base. Of these, 135 individuals were shared between the two solutions, while 10 MAP-OCS selections were dropped and 24 new individuals recruited. The Gini coefficient of contribution weights decreased from 0.483 (MAP-OCS) to 0.460 (CVaR-OCS), a 4.9% reduction indicating a

modest redistribution toward more equitable parental contributions. Evaluated across all 5,000 MCMC scenarios, the gain distribution under CVaR-OCS was significantly less variable than under MAP-OCS (F-test: $p < 0.001$), with no significant difference in mean expected gain (t-test: $p = 0.21$), confirming that CVaR-OCS improves tail-gain security without penalising average performance. A single CVaR-OCS solve at the recommended operating point required 84.0 ± 4.4 s (mean $\pm$ SD, $N = 10$ replicates), compared with 3.9 ± 0.9 s for MAP-OCS, an approximately 22x overhead attributable to the $l = 5,000$ auxiliary variables added by the CVaR formulation (Supplementary Results).

Comparison with ad-hoc robustness-based exclusion

As a reference, we evaluated an intuitive but theoretically ungrounded alternative. By identifying MAP-OCS selections whose robustness scores fell in the bottom quartile ($n = 40$ of 145 selected; mean robustness score = -0.005 vs. $+0.039$ for the retained group), excluding them from the candidate set, and re-running OCS on the remaining individuals. This approach incurred a genetic gain loss of 4.35%, more than six times the cost of CVaR-OCS at the recommended operating point, while paradoxically worsening CVaR₉₅ by 3.7% relative to MAP-OCS. The deterioration reflects a structural limitation of post-hoc exclusion. Removing individuals forces the constrained OCS to concentrate contributions among the remaining high-EBV candidates, increasing sensitivity to posterior uncertainty in exactly those selections. CVaR-OCS avoids this by incorporating uncertainty directly into the optimisation objective, allowing it to simultaneously broaden the selection base. This reduces contribution concentration, and improve tail-gain security, at a fraction of the genetic gain cost.

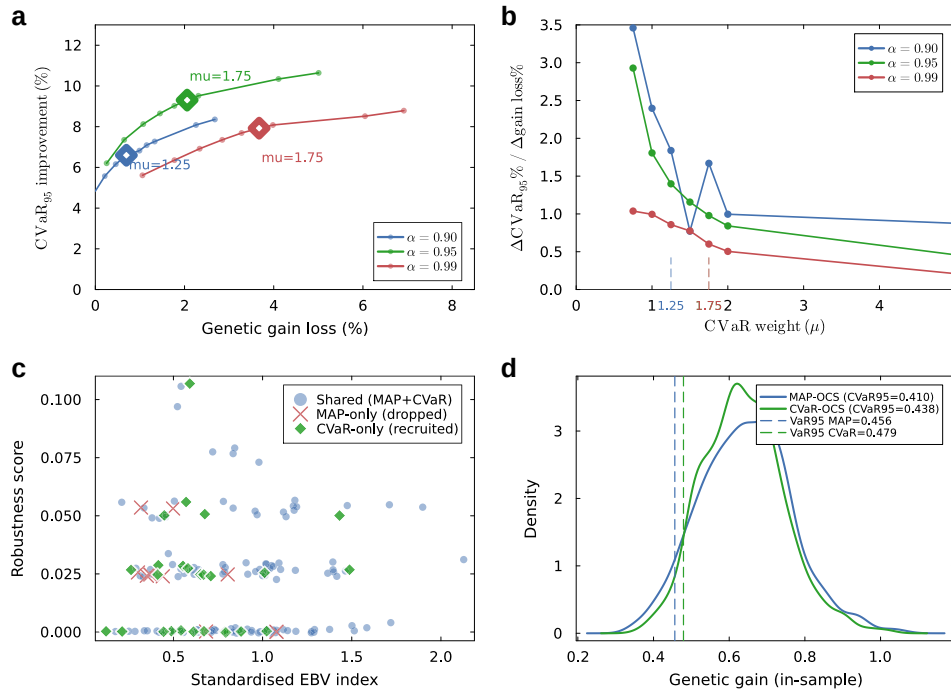


Fig. 4. CVaR-OCS analysis for Norway spruce (*Picea abies*; $n = 1,218$; $\Theta = 0.02$). a) Gain–tail-risk efficiency frontier across three confidence levels ($\alpha \in \{0.90, 0.95, 0.99\}$) and CVaR weights ($\mu \in \{0.50, \dots, 10.00\}$). Each curve shows the trade-off between genetic gain loss (%) and CVaR₉₅ improvement (%) relative to the MAP-OCS baseline. Open diamonds mark the elbow point for each α level, identified by a geometric perpendicular-distance criterion applied to the normalised frontier (CVaR-OCS parameter selection in Materials and Methods). b) Marginal efficiency of CVaR weighting—the ratio of CVaR₉₅ improvement to genetic gain cost between successive μ values. A sharp drop in marginal efficiency confirms the recommended operating point at $\alpha = 0.90$, $\mu = 1.25$ (dashed vertical line); the x-axis is truncated at $\mu = 5$ for clarity. c) Standardised EBV index vs MCMC robustness score for all candidates at the recommended operating point, coloured by selection group: individuals retained by both MAP-OCS and CVaR-OCS (Shared; blue circles), MAP-OCS selections dropped by CVaR-OCS (MAP-only; red crosses), and new individuals recruited by CVaR-OCS (CVaR-only; green diamonds). Robustness score reflects the expected genetic gain loss when an individual is excluded across MCMC iterations. d) Kernel density of genetic gain realised by MAP-OCS (blue) and CVaR-OCS (green) solutions evaluated across all 5,000 MCMC EBV scenarios. Dashed vertical lines mark the 5th-percentile gain (VaR₉₅) for each solution; CVaR₉₅ values (mean gain below the 5th percentile) are given in the legend. CVaR-OCS improves CVaR₉₅ by 6.60% and VaR₉₅ by 5.00% relative to MAP-OCS at a genetic gain cost of 0.70%, confirmed by a significantly reduced gain variance (F-test: $p < 0.001$) with no significant reduction in mean gain (t-test: $p = 0.21$).

Table 3. Summary of CVaR-OCS efficiency frontier elbow points and the MAP-OCS baseline for Norway spruce (*Picea abies*; $n = 1,218$; $\Theta = 0.02$). Elbow points were identified using a geometric perpendicular-distance criterion applied to the normalised gain–CVaR₉₅ efficiency frontier (see Materials and Methods). Gain loss and CVaR₉₅ improvement are expressed as percentages relative to the MAP-OCS baseline. The recommended operating point ($\alpha = 0.90$, $\mu = 1.25$) is shown in bold. Ad-hoc exclusion results are reported for directional comparison only (see text).

Method	α	μ	$\bar{G}$	Gain loss (%)	CVaR ₉₅	CVaR ₉₅ imp. (%)	n_{sel}	Gini
MAP-OCS	—	—	0.6436	—	0.4099	—	145	0.483
CVaR-OCS (elbow)	0.90	1.25	0.6391	−0.70	0.4369	+6.60	159	0.460
CVaR-OCS (elbow)	0.95	1.75	0.6304	−2.06	0.4480	+9.30	159	0.444
CVaR-OCS (elbow)	0.99	1.75	0.6200	−3.67	0.4423	+7.93	170	0.463
Ad-hoc exclusion [†]	—	—	—	−4.35	—	−3.70	118	0.510

[†]Ad-hoc exclusion: bottom-quartile MAP-OCS selections by robustness score ($n = 40$) excluded; OCS re-run on remaining candidates. Absolute gain and CVaR₉₅ values are not directly comparable to CVaR-OCS due to differences in EBV standardisation between analysis pipelines; percentage changes relative to the respective MAP-OCS baseline are reported.

Discussion

Optimum contribution selection (OCS) is a widely used framework for balancing genetic gain and inbreeding in breeding and conservation programs (Meuwissen 1997; Hinrichs et al. 2006; Mullin and Belotti 2015; Wang et al. 2017; Wellmann 2019; Hjortø et al. 2022; El-Kassaby et al. 2024; Waldmann 2025), yet the impact of uncertainty in estimated breeding values (EBVs) on optimized contributions has received comparatively little attention. Here we address this gap through two complementary contributions. First, we introduce CVaR-OCS, a single-solve uncertainty-aware formulation that directly incorporates the joint posterior distribution of

EBVs into the optimization objective via the Conditional Value at Risk (CVaR) risk measure. Second, we use Monte Carlo sampling across MCMC iterations (MCMC-OCS) as a diagnostic tool that exposes the instability of MAP-OCS decisions under EBV uncertainty, and we develop individual-level robustness scores to identify high-risk selections. CVaR-OCS is the principal methodological advance: rather than solving one OCS problem per posterior draw and aggregating the results post hoc, it solves a single quadratic program that explicitly optimizes expected genetic gain subject to a formal tail-risk constraint, providing a principled and computationally efficient path to uncertainty-aware breeding decisions.

The choice of CVaR as the risk measure for OCS rests on solid theoretical foundations from financial portfolio optimization. CVaR, also known as Expected Shortfall, is a coherent risk measure in the sense of Artzner et al. (1999). It satisfies subadditivity, monotonicity, translation invariance, and positive homogeneity. Subadditivity in particular is essential for a diversification-based framework such as OCS, because it guarantees that combining two contribution portfolios cannot increase tail risk beyond the sum of their individual risks. Crucially, CVaR admits a tractable quadratic programming reformulation (Rockafellar and Uryasev 2000, 2002), which is what makes CVaR-OCS practical: the auxiliary-variable representation transforms the otherwise non-smooth tail expectation into a QP with l additional linear constraints, one per MCMC scenario, that can be solved directly with a standard interior-point or ADMM solver such as COSMO (Garstka et al. 2021).

Given the relatively tight relatedness constraint ($\Theta = 0.02$ in the Norway spruce case study), much of the genetic gain captured by OCS derives from exploiting within-family variation. However, when EBV uncertainty is propagated through the posterior distribution, within-family rankings shift dramatically across MCMC iterations compared to the MAP-optimal solution. These instabilities are directly connected to the role of Mendelian segregation variance in OCS (Avendaño et al. 2004; Villanueva et al. 2006; Woolliams et al. 2015; Howard et al. 2018). Avendaño et al. (2004) showed that optimized contributions are allocated according to each candidate's Mendelian sampling term, that is, their superiority over the parental average. Thus, accurate estimation of this within-family additive component is critical for OCS to deliver its theoretical advantages. Howard et al. (2018) confirmed this in a commercial pig population. Our framework makes the consequences of imprecise Mendelian sampling estimates explicit. Individuals whose within-family advantage is poorly estimated receive highly variable contributions across posterior draws, and are precisely the individuals flagged as high-risk by the MCMC robustness scores (Supplementary Material). Strategies to improve within-family precision include broader progeny testing (Kerr et al. 2015), increasing family size and representation (Chu et al. 2019), and enhanced phenotyping protocols (Angidi et al. 2025).

Alongside CVaR-OCS, we developed two complementary descriptive metrics, individual robustness scores and the Gini coefficient of the contribution distribution, for characterising EBV uncertainty at the individual and population level respectively. The population-level metrics connect directly to the theory of long-term genetic contributions formalised by Woolliams et al. (1999) and reviewed by Woolliams et al. (2015): under the genetic contributions framework, the rate of inbreeding is proportional to the sum of squared long-term genetic contributions ($\Delta F \propto \sum r_i^2$), so the distribution of parental contributions determines not only single-generation genetic gain but also the long-run effective population size and inbreeding trajectory (Woolliams et al. 1999; Bijma 2012). A skewed contribution distribution (high Gini, high top-10% concentration) implies a narrow effective gene pool that accelerates genetic drift and inbreeding accumulation across generations, even when the single-generation coancestry constraint Θ is satisfied (Meuwissen et al. 2020). Monitoring the Gini coefficient and concentration ratio alongside CVaR and genetic gain therefore provides a window onto the multi-generational consequences of a given contribution vector, motivating their inclusion as diagnostic complements to the single-generation CVaR objective.

It is important to note that a low individual robustness score can reflect either high EBV uncertainty for that individual across

MCMC iterations, or the availability of near-equivalent genetic substitutes in the candidate pool, or both. The score therefore quantifies replaceability under uncertainty rather than EBV uncertainty in isolation. This distinction is practically relevant in populations with structured family groups, where within-family Mendelian sampling variation means that siblings may share nearly identical EBVs under limited data (Woolliams et al. 1999; Avendaño et al. 2004). An individual with a highly uncertain EBV but no close genetic substitutes may still receive a high robustness score, whereas an individual with a stable EBV but several near-clones in the candidate set may score low. CVaR-OCS does not share this ambiguity, as it optimises directly against the joint posterior uncertainty of the full candidate pool simultaneously.

In Norway spruce, the dual metrics revealed a notable dissociation. Individual robustness improved after post-hoc exclusion of high-risk selections, but contribution concentration increased (Gini +5.6%, top-10% concentration +12.9%), indicating that population-level diversification did not co-move with individual-level stability. The mechanism is structural by construction, as OCS concentrates contributions on the highest-EBV individuals subject to the coancestry constraint (Meuwissen 1997). Removing candidates forces the optimiser to concentrate even more heavily among the remaining elite, reintroducing a different form of fragility elsewhere in the portfolio and, from the perspective of long-term contributions (Woolliams et al. 1999, 2015), narrowing the effective gene pool relative to the MAP-OCS baseline. This contrasts with the QTL-MAS results, where CVaR-OCS simultaneously broadened the selection base, reduced contribution concentration, and improved tail-gain security, i.e. a concordance that likely reflects the less tight coancestry constraint and larger candidate pool in that dataset, which give the optimiser more room to redistribute contributions without sacrificing gain (Gautason et al. 2023). The divergent behaviour between datasets underscores that individual robustness and portfolio concentration are not interchangeable metrics and that post-hoc exclusion cannot in general guarantee improvement on both dimensions simultaneously (Woolliams et al. 2015; Meuwissen et al. 2020). Notably, the robustness scores also revealed that several high-contributing individuals under MAP-OCS were simultaneously classified as high-risk (Supplementary Fig. S11b), illustrating that a large MAP contribution does not imply stability under posterior uncertainty: an individual can dominate the MAP solution because their point-estimate EBV is high relative to the coancestry constraint, while their full posterior distribution overlaps substantially with competing candidates, making them replaceable across many MCMC iterations. This is the class of selections that both robustness scores and CVaR-OCS are designed to identify and penalise respectively.

The bottom-quartile threshold used to classify high-risk selections is illustrative rather than prescriptive. In practice, the threshold could be guided by: (i) a tolerable budget of genetic gain loss, framed within the classical gain-inbreeding trade-off of OCS (Meuwissen 1997); (ii) a natural break in the robustness score distribution, which often separates a small cluster of markedly low-scoring individuals from the remaining candidates; or (iii) the CVaR-OCS efficiency frontier itself, which provides a principled and continuous parameterisation of the robustness-gain trade-off through α and μ without requiring an arbitrary exclusion threshold. We recommend CVaR-OCS as the primary decision-making tool, with robustness scores serving as complementary diagnostics to identify which MAP-OCS selections are most exposed to posterior uncertainty. Further simulation work across multiple generations is needed to establish how the Gini

coefficient and robustness scores behave over time and which monitoring statistics are most informative for long-term operational decision-making (Woolliams et al. 2015; Meuwissen et al. 2020).

The computational demands of the three framework components differ substantially and should guide implementation choices in practice. A single MAP-OCS solve is fast (3.9 ± 0.9 s on the Norway spruce genomic dataset), and CVaR-OCS adds a manageable overhead (84.0 ± 4.4 s; approximately $22\times$ MAP-OCS) arising from the l auxiliary variables introduced by the CVaR reformulation. This is a fixed cost that does not grow with repeated optimisation across scenarios, since CVaR-OCS solves only once regardless of chain length. We consider this overhead justified by the gain in robustness, the elimination of the arbitrary exclusion threshold, and the principled probabilistic interpretation of the CVaR objective; for breeding programs already investing in MCMC genetic evaluation, the additional computation is modest relative to the evaluation step itself. Robustness score calculation is substantially more demanding, scaling as $n_{\text{sel}} \times n_{\text{MCMC}}$ sequential OCS solves. In the Norway spruce scenario this amounted to approximately 1,133 min sequentially, reduced to around 57 minutes on 20 cores after subsampling to 200 MCMC draws (see Results). For large populations or dense genomic relationship matrices, exhaustive robustness score computation will often be impractical, and we recommend restricting it to situations where populations are small, pedigree-based matrices are available (reducing single-solve time by 4–18 fold relative to $\mathbf{G}$; Supplementary Table S9), or where targeted diagnostic information about specific high-value candidates is operationally warranted. For routine uncertainty-aware OCS, CVaR-OCS is the recommended approach precisely because it delivers the robustness benefits of the full MCMC posterior in a single computationally efficient solve.

A recurring practical question in genomic OCS is whether the relationship matrix used in the coancestry constraint must be the same as the one used in the genetic evaluation step to infer breeding values, and if not, which matrix is preferable in each role. In principle, these two uses serve distinct purposes. The evaluation matrix determines the statistical model for decomposing phenotypic variance and thus affects EBV accuracy, while the OCS matrix controls which notion of relatedness is penalised in the coancestry constraint and therefore shapes the long-term inbreeding trajectory. There is no theoretical requirement that these two matrices be identical (Meuwissen et al. 2020), and empirical evidence on their relative merits in the OCS step remains mixed. Henryon et al. (2019) found that pedigree-based OCS (POCS) realised 14–34% more true genetic gain than genomic OCS (GOCS) under ABLUP, and 1.5–5.7% more under GBLUP, when both were constrained to the same rate of true inbreeding on the IBD scale. They attributed this to GOCS unnecessarily penalising allele frequency changes at markers driven by genetic drift, thereby restricting beneficial selection at QTL. By contrast, Gautason et al. (2023) found that when the number of sires is fixed and each approach is constrained on its own kinship scale, genomic OCS using VanRaden's method 1 with base-population allele frequencies realised more genetic gain for a given level of kinship and inbreeding than pedigree-based OCS, because genomic relationships account for Mendelian sampling variation within families that pedigree relationships cannot resolve. The key difference between these two results is the comparison basis. Constraining POCS and GOCS to the same inbreeding rate on the same scale favours pedigree relationships, whereas allowing each approach to operate on its own scale and fixing operational

constraints such as number of sires favours genomic relationships. Applying the same threshold θ to $\mathbf{G}$ and $\mathbf{A}$ implies different effective inbreeding targets because $\mathbf{G}_{ij}$ includes a baseline IBS component reflecting founder allele frequencies, whereas $\mathbf{A}_{ij}$ measures IBD relative to a defined pedigree base (Hill and Weir 2011). In our Norway spruce analyses, this effect was modest (baseline average coancestry 0.012 genomic vs. 0.014 pedigree), but would become more consequential at tighter constraints or in more inbred populations, and practitioners should calibrate θ to the matrix being used rather than applying a universal threshold across matrix types.

Our work builds on two key recent contributions to uncertainty-aware OCS. Pocrnic et al. (2023) demonstrated the value of propagating posterior EBV distributions through OCS using a simulated population, providing the conceptual foundation for the MCMC-OCS diagnostic procedure used here. Fogg et al. (2024) developed robust OCS using standard errors of EBVs, offering a practical solution applicable to frequentist inference settings. CVaR-OCS is complementary to both: it extends the posterior sampling idea of Pocrnic et al. (2023) into a single unified optimization problem, and it subsumes the minimax robust OCS of Fogg et al. (2024) as a limiting case. As the CVaR confidence level $\alpha \rightarrow 1$, CVaR-OCS converges to a worst-case formulation equivalent to SE-based robust OCS (Fogg et al. 2024). A further distinction is that CVaR-OCS propagates the full joint posterior covariance structure of breeding values implicitly through the scenario set, whereas SE-based approaches treat each candidate's uncertainty independently. This joint propagation is particularly relevant in populations with strong family structure or high genomic relatedness among elite candidates. The QTL-MAS simulation results provide an initial empirical validation. CVaR-OCS recovered one additional oracle-optimal individual compared to MAP-OCS and reduced the variance of the gain distribution, consistent with the theoretical expectation that tail-risk protection should move the solution closer to a robustly-optimal portfolio. However, the improvement in Jaccard similarity to the oracle was modest (+11%), and these results derive from a single-generation benchmark with a relatively small candidate population. Multi-generation simulation studies comparing CVaR-OCS and MAP-OCS under known true breeding values are needed to characterize the long-term effects of incorporating EBV uncertainty on genetic gain trajectories and inbreeding accumulation, and to determine whether the performance advantage of CVaR-OCS grows, diminishes, or remains stable across generations. Together with the independent contributions of Fogg et al. (2024) and Pocrnic et al. (2023), these converging efforts underscore the growing recognition that EBV uncertainty is an important and tractable dimension of OCS that the field is only beginning to address systematically.

Several limitations of the present study warrant mention. First, the computational cost of scenario-based CVaR-OCS scales with the number of posterior draws used to represent predictive uncertainty, since each draw enters the optimization as an additional scenario (Supplementary Table S9). For very large candidate sets or dense posterior samples this cost may become limiting, and covariance-based robust formulations, which summarize predictive uncertainty through a single (co)variance structure rather than an explicit scenario set, and may scale more favorably in such settings. Second, our comparison emphasizes the advantage of exploiting the full joint posterior relative to point-estimate OCS and to uncertainty specifications that treat candidates independently; this advantage is most pronounced relative to standard-error- or diagonal-based approaches. Covariance-based Robust OCS (Fogg et al. 2024) can likewise encode correlated prediction

errors, and a direct empirical comparison of the two paradigms under matched conditions, including how risk aversion is controlled through the CVaR confidence level and penalty weight versus an uncertainty-set radius: this is an important direction we leave to future work. Third, our selection of α and μ via the efficiency-frontier elbow, while practical and reproducible, remains a heuristic that does not explicitly encode a breeder's utility function; linking these parameters to economic weights (Hazel 1943; Simianer et al. 2023) or formal risk-preference elicitation (cf. Akdemir and Sánchez 2016) would provide a more principled operating point. Finally, our oracle validation against known true breeding values is confined to a single generation of selection; the consequences of uncertainty-aware contribution allocation for genetic gain and diversity maintenance across repeated cycles of a breeding program remain to be established, and represent a natural target for multi-generation simulation studies.

The magnitude of CVaR-OCS's tail-gain benefit varied substantially across the datasets examined here, from CVaR₉₅ improvements of 0.46–0.70% in the QTL-MAS simulation to 2.5–3.6% in Loblolly pine and 6.60% in Norway spruce. This variation is consistent with the method's underlying rationale: CVaR-OCS derives its advantage over point-estimate selection by exploiting heterogeneity in prediction uncertainty across candidates, and its benefit should scale accordingly. The QTL-MAS population is a simulated, densely connected pedigree with moderate, relatively homogeneous EBV accuracy across candidates; the real forest tree datasets exhibit substantially more heterogeneous uncertainty, arising from unequal trial connectivity (Kerr et al. 2015), variable family sizes (Papin et al. 2024), and multi-site designs (Chen et al. 2018). The comparatively modest QTL-MAS effect therefore represents a conservative test case rather than a limitation of the approach, and the larger benefits observed in Norway spruce and Loblolly pine are likely more representative of the heterogeneous prediction accuracy typical of operational breeding programmes.

In summary, standard optimum contribution selection relies on point estimates of breeding values and thereby discards the predictive uncertainty inherent in genetic evaluation. Against an oracle with access to true breeding values, point-estimate OCS realized only a fraction of the achievable genetic gain, illustrating the cost of treating uncertain EBVs as fixed. We have introduced CVaR-OCS, a single-solve formulation that embeds the full joint posterior distribution of estimated breeding values into the contribution optimization through the Conditional Value at Risk, a coherent risk measure that simultaneously targets expected gain and guards against the poor outcomes that arise when uncertain candidates are over-weighted. Across a simulated dataset with known true breeding values and empirical forest tree data, CVaR-OCS matched the expected gain of point-estimate OCS while improving tail-gain security and diversifying parental contributions, all within a single optimization rather than the many repeated solves needed to characterize solution stability post hoc. Because prediction uncertainty varies substantially among candidates rather than by generation alone, accounting for it directly yields selection decisions that are more robust to the specific evaluation on which they are based. We anticipate that uncertainty-aware contribution selection of this kind will be most valuable in programs where evaluation accuracy is uneven across selection candidates, which is a common situation in tree and plant breeding, where heterogeneous trial environments, variable connectivity, and unbalanced family structure lead to candidate-specific differences in prediction accuracy.

Data availability

Norway spruce phenotypic and genotypic data with recoded IDs are provided here: https://github.com/jonhar97/MCMC_uncertainty_OCS/.

Supplemental material available at [GENETICS](#) online.

Acknowledgments

This work was supported by grant from Föreningen Skogsträdsförädling project ID 23-511. The authors thank the Associate Editor Dr Jeffrey Endelman and three anonymous reviewers for their constructive comments, which substantially improved the manuscript.

Funding

JA was supported by project ID 21414 (Föreningen ID 23-511) funded by Föreningen Skogsträdsförädling.

Conflicts of interest

The authors declare that they have no conflict of interest.

Literature cited

- Ahlinder J, Mullin TJ, Yamashita M. 2013. Using semidefinite programming to optimize unequal deployment of genotypes to a clonal seed orchard. *Tree Genet Genomes* 10:27–34. <https://doi.org/10.1007/s11295-013-0659-z>.
- Akdemir D, Sánchez JI. 2016. Efficient breeding by genomic mating. *Front Genet.* 7:210. <https://doi.org/10.3389/fgene.2016.00210>.
- Amadeu RR, Garcia AAF, Munoz PR, Ferrão LFV. 2023. Aghmatrix: genetic relationship matrices in R. *Bioinformatics*. 39:btad445. <https://doi.org/10.1093/bioinformatics/btad445>.
- Angidi S, Madankar K, Tehseen MM, Bhatla A. 2025. Crops. 5:8. Advanced high-throughput phenotyping techniques for managing abiotic stress in agricultural crops—a comprehensive review. <https://doi.org/10.3390/crops5020008>.
- Artzner P, Delbaen F, Eber JM, Heath D. 1999. Coherent measures of risk. *Math Financ.* 9:203–228. <https://doi.org/10.1111/1467-9965.00068>.
- Avendaño S, Woolliams JA, Villanueva B. 2004. Mendelian sampling terms as a selective advantage in optimum breeding schemes with restrictions on the rate of inbreeding. *Genet Res (Camb)*. 83:55–64. <https://doi.org/10.1017/S0016672303006566>.
- Bijma P. 2012. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the accuracy of predicted phenotypes in selected populations. *J Anim Breed Genet.* 129: 345–358. <https://doi.org/10.1111/j.1439-0388.2012.00991.x>.
- Birge JR, Louveaux F. 2011. *Introduction to Stochastic Programming*. 2nd ed. Springer.
- Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. 2010. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found Trends Mach Learn.* 3:1–122. <https://doi.org/10.1561/22000000016>.
- Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. *Nat Rev Genet.* 10:783–796. <https://doi.org/10.1038/nrg2664>.
- Chen ZQ et al. 2018. Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. *BMC Genom.* 19:1–16. <https://doi.org/10.1186/s12864-018-5256-y>.

- Cheng H, Kizilkaya K, Zeng J, Garrick D, Fernando R. 2018. Genomic prediction from multiple-trait bayesian regression methods using mixture priors. *Genetics*. 209:89–103. <https://doi.org/10.1534/genetics.118.300650>.
- Christensen OF, Madsen P, Nielsen B, Ostensen T, Su G. 2012. Single-step methods for genomic evaluation in pigs. *Animal*. 6:1565–1571. <https://doi.org/10.1017/S1751731112000742>.
- Chu TT, Bastiaansen JW, Berg P, Komen H. 2019. Optimized grouping to increase accuracy of prediction of breeding values based on group records in genomic selection breeding programs. *Genet Sel Evol*. 51:64. <https://doi.org/10.1186/s12711-019-0509-z>.
- Clark SA, Kinghorn BP, Hickey JM, van der Werf JH. 2013. The effect of genomic information on optimal contribution selection in live-stock breeding programs. *Genet Sel Evol*. 45:44. <https://doi.org/10.1186/1297-9686-45-44>.
- Cowling WA et al. 2023. Optimal contribution selection improves the rate of genetic gain in grain yield and yield stability in spring canola in australia and canada. *Plants*. 12:383. <https://doi.org/10.3390/plants12020383>.
- Derks MFL, Steensma M. 2021. Balancing selection for deleterious alleles in livestock. *Front Genet*. 12:761728. <https://doi.org/10.3389/fgene.2021.761728>.
- Dunning I, Huchette J, Lubin M. 2017. Jump: a modeling language for mathematical optimization. *SIAM Rev*. 59:295–320. <https://doi.org/10.1137/15M1020575>.
- El-Kassaby YA, Cappa EP, Chen C, Ratcliffe B, Porth IM. 2024. Efficient genomics-based “end-to-end” selective tree breeding framework. *Heredity* (Edinb). 132:98–105. <https://doi.org/10.1038/s41437-023-00667-w>.
- Endelman JB. 2011. Ridge regression and other kernels for genomic selection with r package rrblup. *Plant Genome*. 4:250–255. <https://doi.org/10.3835/plantgenome2011.08.0024>.
- Fogg J, Ortiz J, Pocrnić I, Hall JAJ, Gorjanc G. 2024. Robust optimal contribution selection [preprint]. *arXiv preprint arXiv:2412.02888*. <https://doi.org/10.48550/arXiv.2412.02888>.
- Garstka M, Cannon M, Goulart P. 2021. Cosmo: a conic operator splitting method for large convex problems. *Journal of Optimization Theory and Applications*. 190:779–810. <https://doi.org/10.1007/s10957-021-01896-x>.
- Gautason E, Sahana G, Guldbrandtsen B, Berg P. 2023. Impact of kinship matrices on genetic gain and inbreeding with optimum contribution selection in a genomic dairy cattle breeding program. *Genet Sel Evol*. 55:48. <https://doi.org/10.1186/s12711-023-00826-x>.
- Goddard ME. 1994. Selection indices for non-linear profit functions. *Theor Appl Genet*. 64:339–344. <https://doi.org/10.1007/BF00274177>.
- Hallander J, Waldmann P. 2009. Optimum contribution selection in large general tree breeding populations with an application to Scots pine. *Theor Appl Genet*. 118:1133–1142. <https://doi.org/10.1007/s00122-009-0968-7>.
- Hamilton MG. 2020. Optimal contribution selection in highly fecund species with overlapping generations. *J Hered*. 111:646–651. <https://doi.org/10.1093/jhered/esaa051>.
- Hazel LN. 1943. The genetic basis for constructing selection indexes. *Genetics*. 28:476–490. <https://doi.org/10.1093/genetics/28.6.476>.
- Henderson CR. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*. 31:423–447. <https://doi.org/10.2307/2529430>.
- Henryon M et al. 2019. Pedigree relationships to control inbreeding in optimum-contribution selection realise more genetic gain than genomic relationships. *Genet Sel Evol*. 51:39. <https://doi.org/10.1186/s12711-019-0475-5>.
- Hershberger J, Morales N, Simoes C et al. 2021. Making waves in breedbase: an integrated spectral data storage and analysis pipeline for plant breeding programs. *Plant Phenome J*. 4:e20012. <https://doi.org/10.1002/ppj2.20012>.
- Hill WG, Weir BS. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet Res (Camb)*. 93:47–64. <https://doi.org/10.1017/S0016672310000480>.
- Hinrichs D, Wetten M, Meuwissen THE. 2006. An algorithm to compute optimal genetic contributions in selection programs with large numbers of candidates. *J Anim Sci*. 84:3212–3218. <https://doi.org/10.2527/jas.2006-145>.
- Hjortø L, Andersen T, Kargo M, Sørensen AC. 2022. Breeding schemes with optimum-contribution selection or truncation selection for beef cattle destined for use on dairy females. *J Dairy Sci*. 105:4314–4323. <https://doi.org/10.3168/jds.2021-21258>.
- Howard DM, Pong-Wong R, Knap PW, Kremer VD, Woolliams JA. 2018. Selective advantage of implementing optimal contributions selection and timescales for the convergence of long-term genetic contributions. *Genet Sel Evol*. 50:24. <https://doi.org/10.1186/s12711-018-0392-z>.
- Isik F, McKeand SE. 2019. Fourth cycle breeding and testing strategy for pinus taeda in the nc state university cooperative tree improvement program. *Tree Genet Genomes*. 15. 70. <https://doi.org/10.1007/s11295-019-1377-y>.
- Kerr RJ, Dutkowski GW, Jansson G, Persson T, Westin J. 2015. Connectedness among test series in mixed linear models of genetic evaluation for forest trees. *Tree Genet Genomes*. 11:67. <https://doi.org/10.1007/s11295-015-0887-5>.
- Klein Haneveld WK, van der Vlerk MH, Romeijnders W. 2015. *Stochastic Programming*. Springer.
- Legarra A, Aguilar I, Misztal I. 2009. A relationship matrix including full pedigree and genomic information. *J Dairy Sci*. 92:4656–4663. <https://doi.org/10.3168/jds.2009-2061>.
- Lynch M, Walsh B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates.
- Markowitz H. 1952. Portfolio selection. *J Finance*. 7:77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>.
- Markowitz HM. 1959. *Portfolio selection: efficient diversification of investments*. John Wiley & Sons.
- McRae T et al. 2003. TREPLAN - a genetic evaluation system for forest trees. 27th Southern Forest Tree Improvement Conference; 2003 Jun 25–27; Stillwater, OK, USA.
- Meuwissen TH, Sonesson AK, Gebregiorgis G, Woolliams JA. 2020. Management of genetic diversity in the era of genomics. *Front Genet*. 11:880. <https://doi.org/10.3389/fgene.2020.00880>.
- Meuwissen THE. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *J Anim Sci*. 75:934–940. <https://doi.org/10.2527/1997.754934x>.
- Mogensen PK, Riseth AN. 2018. Optim: a mathematical optimization package for Julia. *J Open Source Softw*. 3:615. <https://doi.org/10.21105/joss.00615>.
- Mullin TJ, Belotti P. 2015. Using branch-and-bound algorithms to optimize selection of a fixed-size breeding population under a relatedness constraint. *Tree Genet Genomes*. 12:4. <https://doi.org/10.1007/s11295-015-0961-z>.
- Papin V, Gorjanc G, Pocrnić I, Bouffier L, Sanchez L. 2024. Unlocking genome-based prediction and selection in conifers: the key role of within-family prediction accuracy. *Ann For Sci*. 81:52. <https://doi.org/10.1186/s13595-024-01269-0>.
- Pocrnić I, Ortiz J, Gorjanc G. 2023. Evaluating the impact of uncertainty in estimated breeding values on optimal contribution selection. *Proceedings of the 74th Annual Meeting of the*

- European Federation of Animal Science (EAAP); Lyon, France. Wageningen Academic Publishers.
- Pong-Wong R, Woolliams JA. 2007. Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming. *Genet Sel Evol.* 39:3. <https://doi.org/10.1186/1297-9686-39-1-3>.
- Rasmussen CE, Williams CK. 2006. *Gaussian Processes for Machine Learning*. MIT Press.
- Resende JF et al. 2012. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics.* 190: 1503–1510. <https://doi.org/10.1534/genetics.111.137026>.
- Rockafellar RT, Uryasev S. 2000. Optimization of conditional value-at-risk. *J Risk.* 2:21–41. <https://doi.org/10.21314/JOR.2000.038>.
- Rockafellar RT, Uryasev S. 2002. Conditional value-at-risk for general loss distributions. *J Bank Finance.* 26:1443–1471. [https://doi.org/10.1016/S0378-4266\(02\)00271-6](https://doi.org/10.1016/S0378-4266(02)00271-6).
- Satopaa V, Albrecht J, Irwin D, Raghavan B. 2011. Finding a “kneedle” in a haystack: detecting knee points in system behavior. *Proceedings of the 31st International Conference on Distributed Computing Systems Workshops (ICDCSW)*; p. 166–171.
- Shapiro A, Dentcheva D, Ruszczyński A. 2009. *Lectures on stochastic programming: modeling and theory*. SIAM.
- Simianer H et al. 2023. How economic weights translate into genetic and phenotypic progress, and vice versa. *Genet Sel Evol.* 55: 38. <https://doi.org/10.1186/s12711-023-00807-0>.
- Sorensen D, Gianola D. 2002. *Likelihood, Bayesian, and MCMC methods in quantitative genetics*. Springer.
- Szydlowski M, Paczyńska P. 2011. QTLMAS 2010: simulated dataset. *BMC Proc.* 5:S3. <https://doi.org/10.1186/1753-6561-5-S3-S3>.
- Tao L et al. 2025. The effects of runs-of-homozygosity on pig domestication and breeding. *BMC Genom.* 26:6. <https://doi.org/10.1186/s12864-024-11189-y>.
- Tchounke B, Sanchez L, Bell JM, Cros D. 2023. Mate selection: a useful approach to maximize genetic gain and control inbreeding in genomic and conventional oil palm (*Elaeis guineensis* Jacq.) hybrid breeding. *PLoS Comput Biol.* 19:e1010290. <https://doi.org/10.1371/journal.pcbi.1010290>.
- van der Wilk M et al. 2020. A framework for interdomain and multi-output Gaussian processes [preprint]. arXiv:2003.01115. <https://doi.org/10.48550/arXiv.2003.01115>.
- VanRaden PM. 2008. Efficient methods to compute genomic predictions. *J Dairy Sci.* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- Villanueva B, Avendaño S, Woolliams JA. 2006. Prediction of genetic gain from quadratic optimisation with constrained rates of inbreeding. *Genet Sel Evol.* 38:127–146. <https://doi.org/10.1186/1297-9686-38-2-127>.
- Waldmann P. 2025. Genomic optimum contribution selection and mate allocation using jump. *Bioinform Adv.* 5:vbaf259. <https://doi.org/10.1093/bioadv/vbaf259>.
- Wang Y, Bennewitz J, Wellmann R. 2017. Novel optimum contribution selection methods accounting for conflicting objectives in breeding programs for livestock breeds with historical migration. *Genet Sel Evol.* 49:45. <https://doi.org/10.1186/s12711-017-0320-7>.
- Weigel KA, Rekaya R, Zwald NR, Fikse WF. 2001. International genetic evaluation of dairy sires using a multiple-trait model with individual animal performance records. *J Dairy Sci.* 84:2789–2795. [https://doi.org/10.3168/jds.S0022-0302\(01\)74734-0](https://doi.org/10.3168/jds.S0022-0302(01)74734-0).
- Wellmann R. 2019. Optimum contribution selection for animal breeding and conservation: the R package optisel. *BMC Bioinform.* 20:25. <https://doi.org/10.1186/s12859-018-2450-5>.
- Woolliams JA, Berg P, Dagnachew BS, Meuwissen TH. 2015. Genetic contributions and their optimization. *J Anim Breed Genet.* 132: 89–99. <https://doi.org/10.1111/jbg.12148>.
- Woolliams JA, Bijma P, Villanueva B. 1999. Expected genetic contributions and their impact on gene flow and genetic gain. *Genetics.* 153:1009–1020. <https://doi.org/10.1093/genetics/153.2.1009>.
- Yamashita M, Mullin TJ, Safarina S. 2018. An efficient second-order cone programming approach for optimal selection in tree breeding. *Optim Lett.* 12:1683–1697. <https://doi.org/10.1007/s11590-018-1229-y>.

Editor: J. Endelman