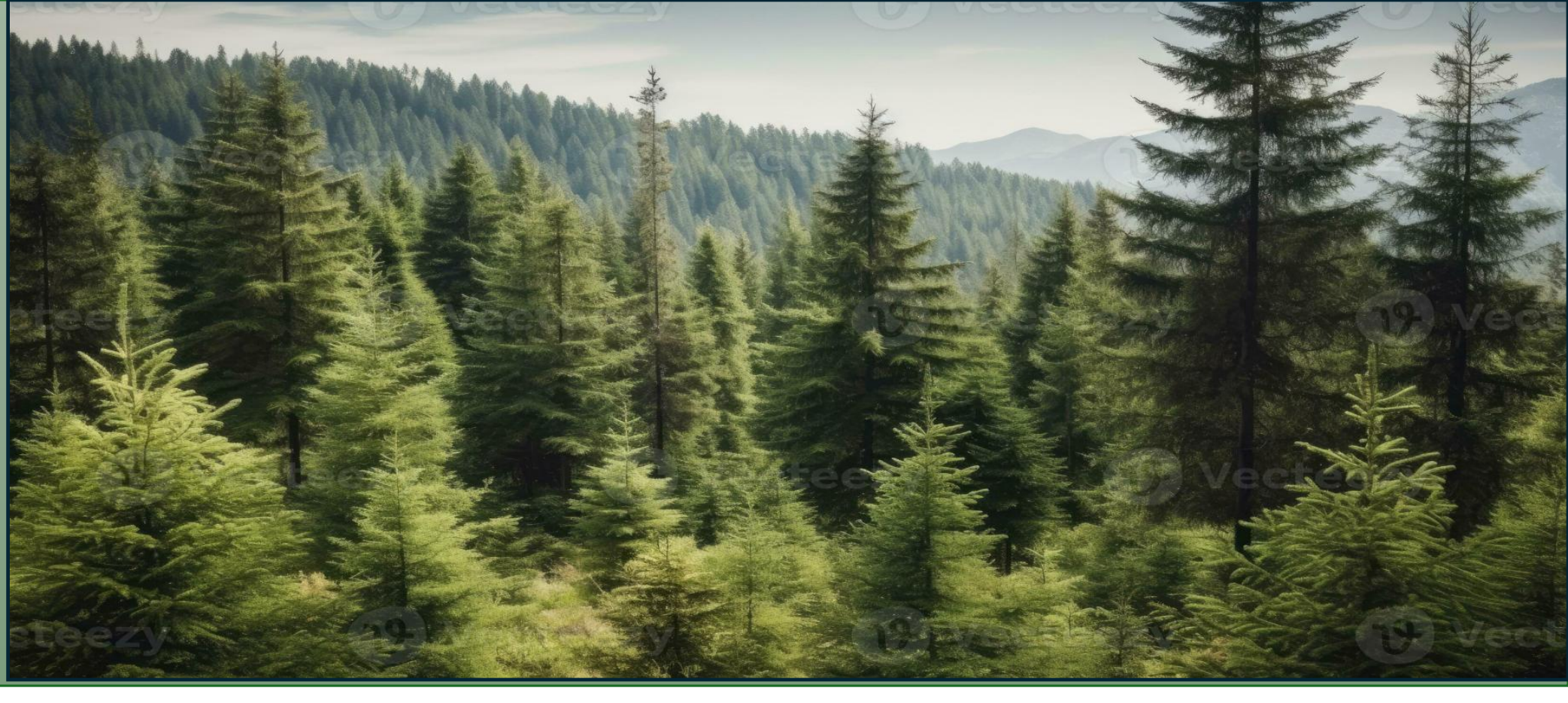
Beyond Endpoint Measurements: Modeling Phenotypic Trajectories for Climate-Informed Genomic Prediction

Soumodip Pal[1], Lan Zhu[1], Barb R. Thomas[2], the RES-FOR Project and Charles Chen[3]

[1] Department of Statistics, [3]Department of Biochemistry and Molecular Biology ,Oklahoma State University, Stillwater, OK, USA.
[2]Department of Renewable Resource, Faculty of Agricultural, Life & Environmental Sciences, University of Alberta, Edmonton, AB, Canada.

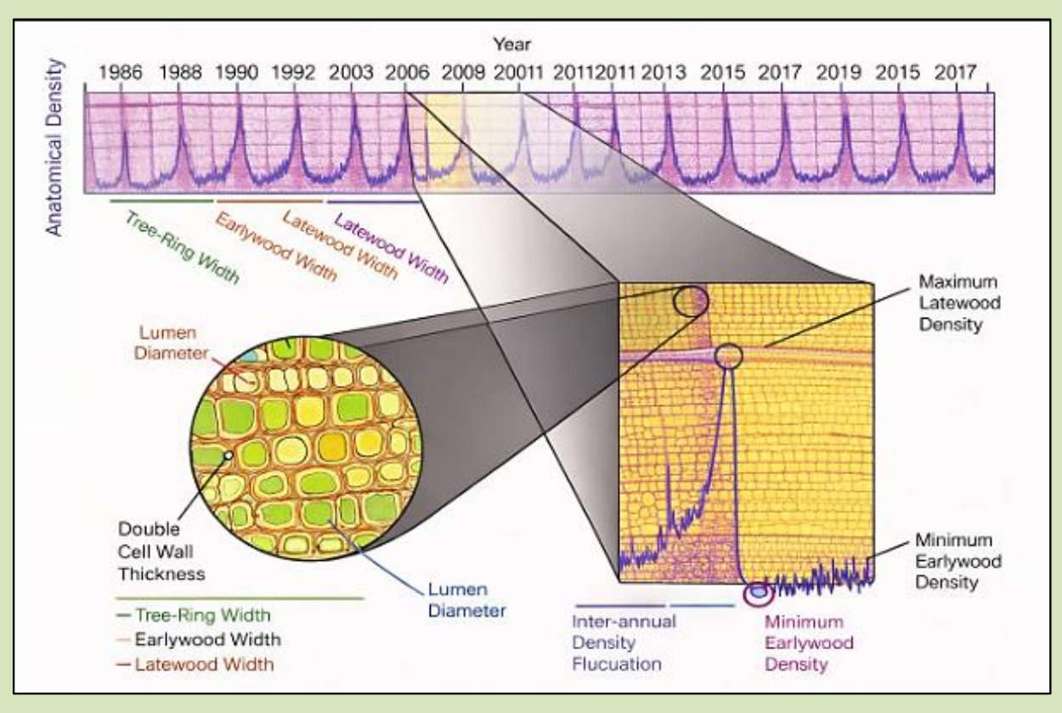


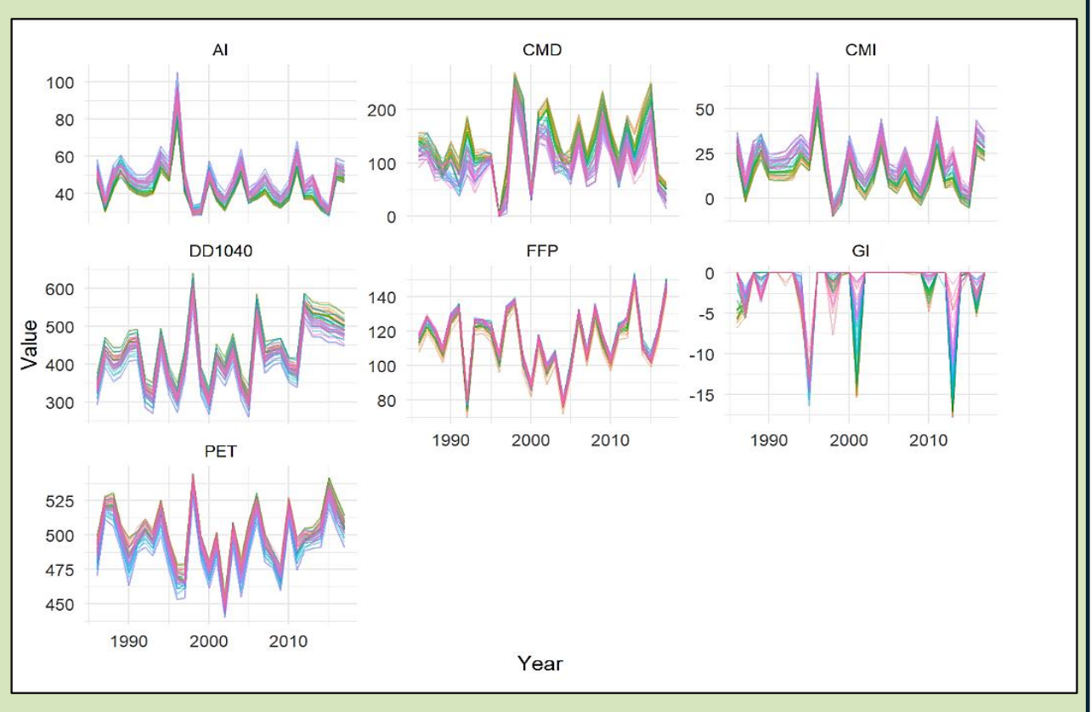
Introduction

- Biological growth is inherently a *dynamic, **nonlinear trajectory*** shaped by continuous gene–environment interaction across plant and animal systems.
- Conventional genomic prediction aggregates trajectories into ***static endpoint summaries***.
- Static prediction lack in explaining ***how and when*** traits unfold through time and respond to environmental variation.
- Despite advances in G×E modeling, longitudinal growth, climate covariates, and genetic structure are rarely unified within a ***single inferential framework***.
- This gap limits environment-based prediction in breeding, conservation, and assisted adaptation.

Objective

- Develop a **framework** that operates on **nonlinear growth trajectories**, enabling biologically interpretable inference and uncertainty-aware prediction across time and environments.
- **Broadly applicable to high-throughput phenotyping** in plant and animal systems.
- we demonstrate the model's utility through growth of **tree-rings** which serve as a natural archive of climate impacts on tree growth.
- Model **climate based genomic prediction** which is critical for climate-sensitive seed transfer, reforestation planning, and long-term forest resilience.





Data

We used interior lodgepole pine *(Pinus contorta var. latifolia Douglas*) trees, sampled throughout the growth period from 1986-2017, from half-sib families across four locations (Judy creek, Virginia, Swan lake, Timeau) in central Alberta, Canada, to assess genetic and environmental regulators of growth. As for the genotype data we considered around 46k SNPs across 200 scaffolds to compute the kinship matrix (G). By using the coordinates of the sites, we collected the historical and future annual climate data from *ClimateNA* web version.

Statistical Modelling

In this study, we addressed the knowledge gap using a hierarchical Bayesian framework that models tree-ring data as growth trajectories, capturing nonlinear growth curve through a ***Gompertz function (S-shaped)***

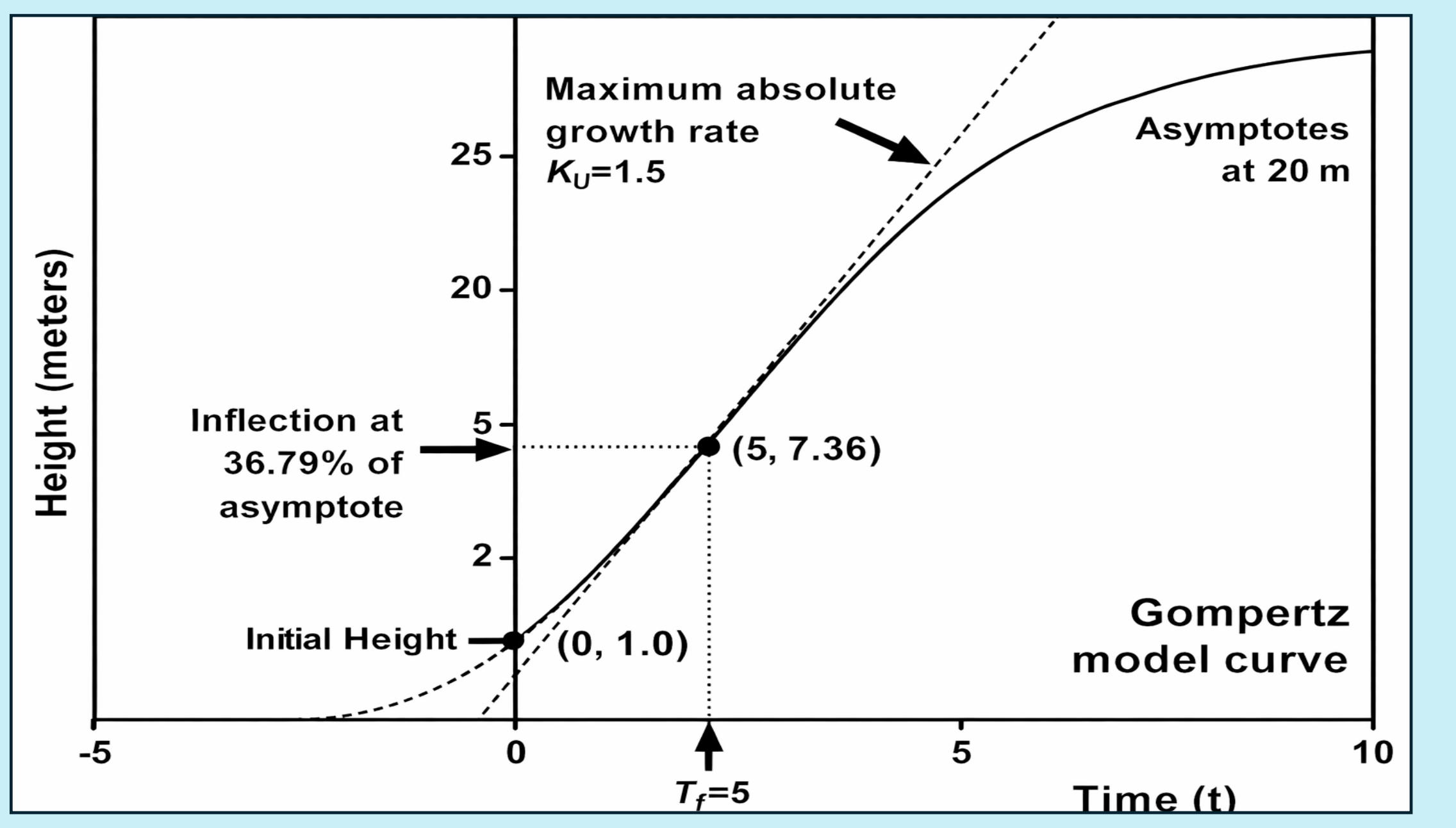
$$R = \boldsymbol{A}\, exp[-exp\{-\boldsymbol{K}(t - \boldsymbol{\Psi})\}] \qquad \ldots\ldots.(1)$$

Each Gompertz parameter—representing maximum growing capacity (**A**), growth rate (**K**), and inflection point (**Ψ**) — is further decomposed into key climate-resilience parameters, such as site effect, overall additive genetic effects, and climate sensitivity throughout growth period.

for each,

$$\Theta \in \{A, K, \Psi\}, \quad log(\Theta) = \boldsymbol{\mu} + \boldsymbol{\alpha} + \mathbf{P} + \mathbf{S} + \mathbf{C} \qquad \ldots\ldots.(2)$$

- ✓ This hierarchical decomposition separates genetic (**α**), permanent(**P**), site (**S**)and environmental(**C**) effects, while accommodating uncertainty through fully Bayesian inference, revealing how climate variability shapes growth potential.
- ✓ This approach enables probabilistic forecasting of growth trajectories for each genotype in unobserved future environments.
- ✓ We further extend posterior predictive simulations to estimate tree-ring growth parameters under the projected climate conditions.



"Instead of predicting growth year by year, we are interested in predicting how each individual grows, and how climate shifts that trajectory."

Modelling Architecture

The genetic components for each tree are collected into a six-dimensional vector (refer (1) and (2)) with a kinship-coupled multivariate normal prior such that

$$vec(G_{mat}) \sim N(0, \Sigma g \otimes G)$$

This structure captures both genetic correlations among growth parameters and genetically structured responses to climate.

We chose weakly informative priors for the effects.

If ϕ denotes all hierarchical components. The posterior density is

$$p(\phi \mid R) \propto p(R \mid \Theta, \sigma^2) p(\Theta \mid \phi) p(\phi) \qquad \ldots\ldots\ldots..(3)$$

where Θ is deterministically computed from ϕ.

Because the Gompertz curve induces a nonlinear likelihood, the posterior has no closed-form expression. We therefore use Markov chain Monte Carlo methods, with Metropolis–Hastings (MH) design.

Prediction Capacity

Along with the future climate-based prediction on the existing sites and known genotypes . Our model is also capable to predict growth trajectories for new genotypes and new sites.

- For a ***new sample***(genotype), draw the posterior conditional on kinship.
- For a ***new site***, draw the site effect corresponding to the new site or a posterior predictive value if the site is unobserved.

This way the result is a full posterior distribution over future growth trajectories, enabling quantification of uncertainty in both genetic and climatic components.

Overall, the outcome of this study provides actionable insights for climate-adaptive forestry strategies, supporting proactive genotype selection for orchard parents, reforestation planning, and long-term forest resilience.

Simulation Study

We designed our simulation study based on this following key questions:

- Can we recover genetic growth parameters?
- Can we separate climate effects from time trends?
- Can we distinguish genetic vs permanent environmental effects?
- What sample sizes are needed?

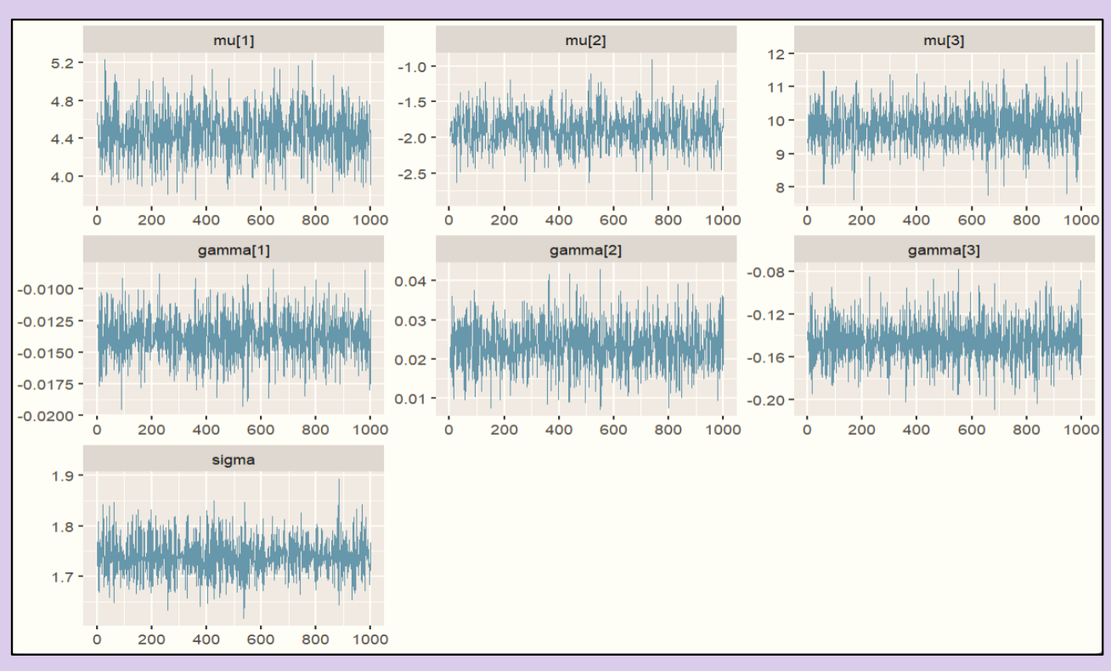
Simulation Results: Model Validation and Identifiability

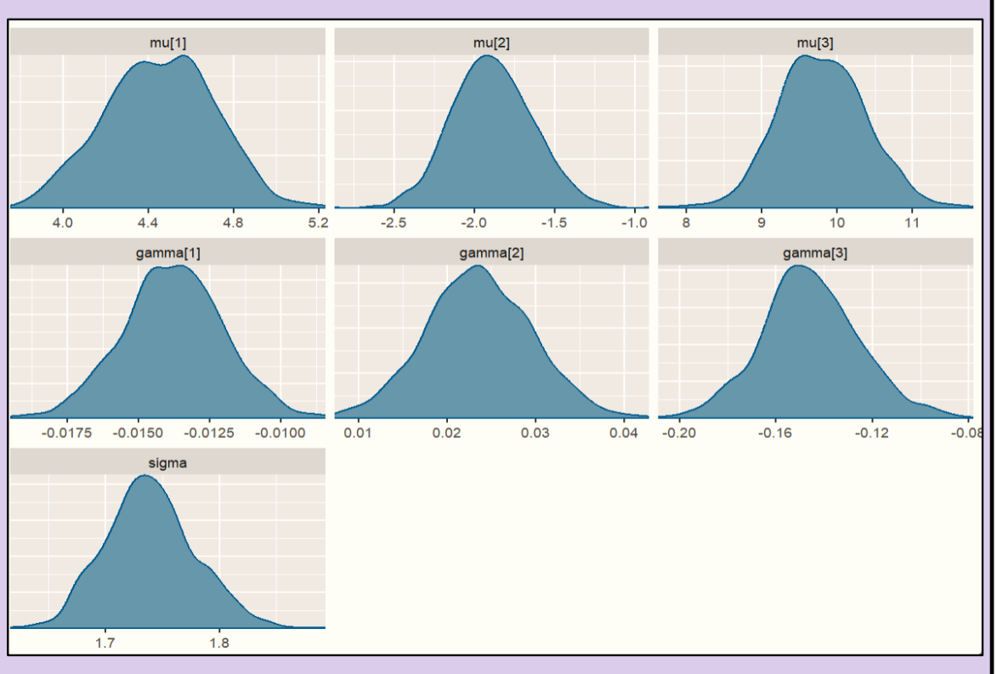
Simulation experiments calibrated to the real lodgepole pine data demonstrate:

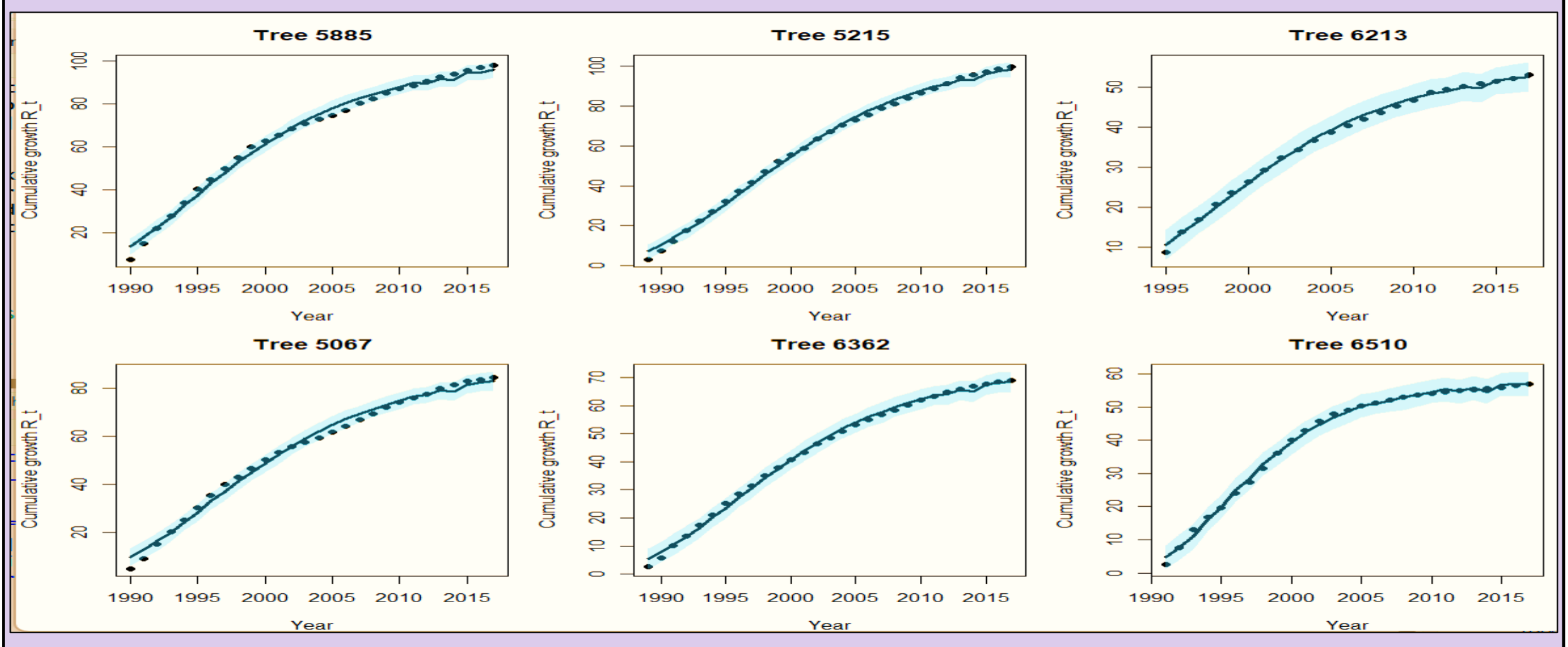
- ☐ The proposed framework **reliably recovers** genotype-specific growth parameters and climate sensitivities under realistic sampling conditions.
- ☐ When *climate variation is not fully confounded with time*, genetic effects and climate responses are **jointly identifiable, validating the use of this prediction.**
- ☐ When *climate variation is confounded with time*, uncertainty in climate sensitivity appropriately increases rather than producing spurious effects, **highlighting the model's robustness and conservative behavior.**

Results and Findings

- MCMC shows **stable exploration** and **good mixing** for global parameters—strong evidence that the hierarchical trajectory model is computationally feasible for the full pipeline.
- Current results revealed a **significant trade-off** ($r = -0.6$) between growing capacity (A) and growth rate (K).
- **Warmer locations** has modestly **accelerated early-season growth.** (A ≈ 82)
- **Colder locations** are likely to experience a **reduction in maximum growing capacity.** (↓A)
- Most between-tree heterogeneity concentrates in **timing (Ψ)** rather than just size or rate—supporting our claim that **climate resilience is often "when growth happens," not only "how much."**
- Using **No U-Turn Sampler (NUTS)** we achieved higher mean acceptance rate ≈ 0.97 and tight global parameter intervals ***enabling stable estimation*** of climate-driven trajectory shifts.
- Our posterior predictive simulations reproduced individual trajectories accurately: observed growth of ring width for representative trees fell within the **95% posterior intervals**, and residuals showed **no systematic bias**.







Limitation and Future Work

Although the hierarchical Bayesian framework provides rich uncertainty quantification and biologically interpretable inference, full MCMC inference remains **computationally intensive**, particularly when estimating high-dimensional genetic covariance structures and genotype × climate interactions across thousands of trees.

1. Ongoing and planned improvements include more aggressive reparameterization and non-centering strategies.
2. Partial marginalization of random effects, and block-wise updating schemes to improve effective sample size per second.

References

Cargnin, M., Gianola, D., & Rosa, G. J. M. (2015). Genomic prediction of genotype × environment interactions with implications for genomic selection. *Genetics Selection Evolution*, 47, 63.
Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Pritchett, J. G., Eyherabide, G. H., … & Burgueño, J. (2014). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science*, 19(10), 592-601
Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46(3), 673-687.
Narinc, D., Aksoy, T., Karabag, K., & Firat, M. Z. (2009). Growth curve parameters and correlations between growth curve parameters in Turkish quail: application of Gompertz, Logistic and Richards models. *Archiv für Geflügelkunde*, 73(3), 169-177.
Piray, H., Dezfouli, A., Dezfouli, M., Heskes, T., & Daw, N. D. (2021). Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. *PLoS Computational Biology*, 15(7), e1007043.
Vaghi, C., Rodallec, A., Fanciullino, R., Ciccolini, J., Mochel, J. P., Mastri, M., … & Benzekry, S. (2020). Population modeling of tumor growth curves and the reduced Gompertz model improve prediction of the age of experimental tumors. *PLOS Computational Biology*, 16(2), e1007178

Acknowledgement

We acknowledge cash funding for this research from Alberta Innovates, the Forest Resource Improvement Program with contributions from West Fraser (Hinton, Sundre, Blue Ridge, North Central Woodlands, Manning, and Cochrane), Weyerhaeuser Canada (Drayton Valley and Grande Prairie), Canadian Forest Products (Grande Prairie and Whitecourt), Tolko, Alberta Newsprint Company, and Northland Forest Products. Additional funding was provided by Project Learning Tree Canada.

In-kind funding has been provided by Alberta Forestry and Parks, Northland Forest Products, the University of Alberta, Oklahoma State University, and the Huallen Seed Orchard Company.

Please reach out to me at soumodip.pal@okstate.edu or Chen's lab at charles.chen@okstate.edu for any queries