



A look at computations for multivariate single-step genomic evaluation fitting the ‘hybrid model’

Karin Meyer

Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351

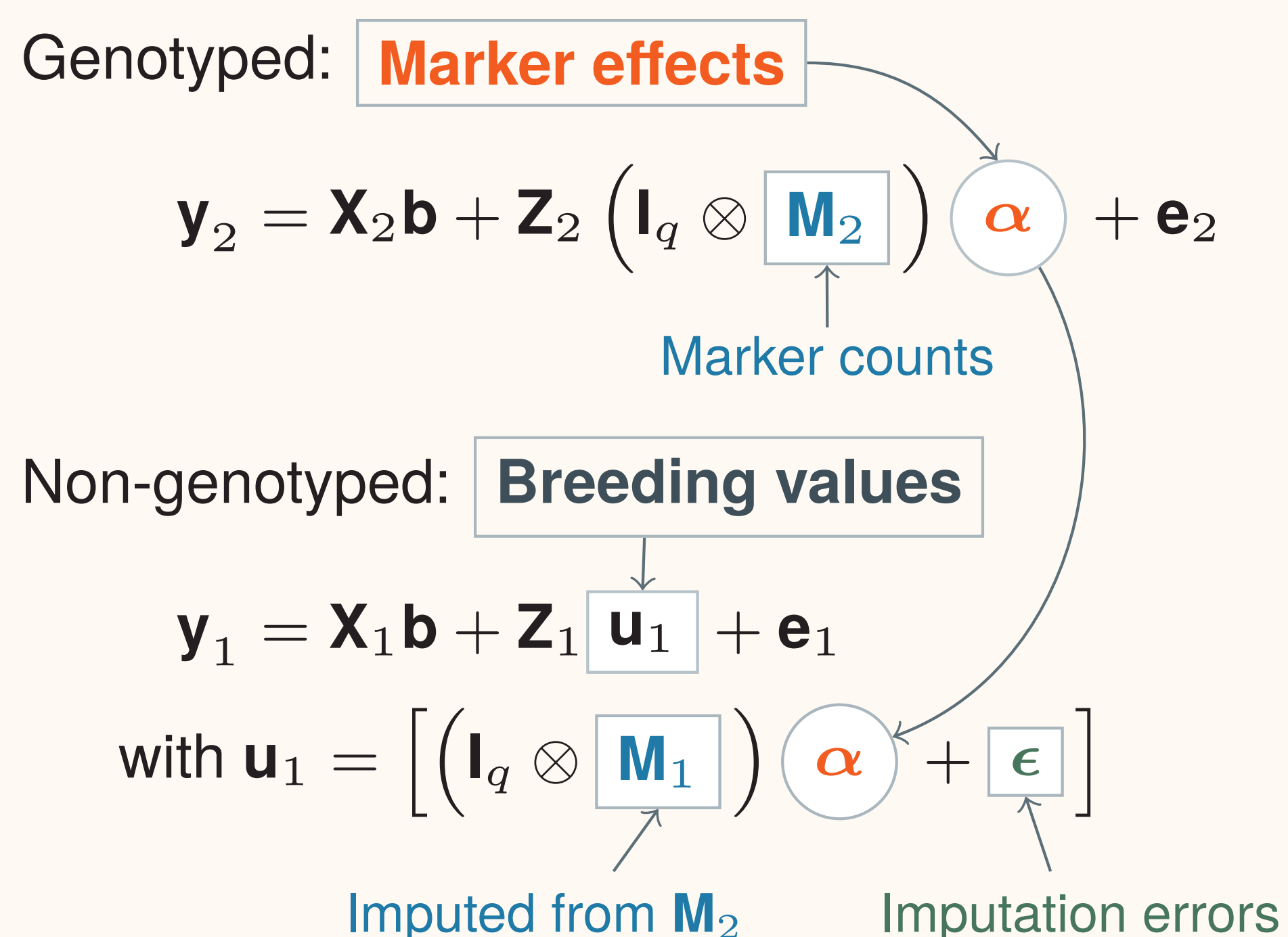
Objectives

- Examine ‘Hybrid Model’ for single-step genomic BLUP

Fernando et al. 2016, GSE 48:96

- Implementation issues
- Computational demands
- Ease of parallel processing
- Many traits

Hybrid model



Simulated data

- 1.5 million animals, 3 generations
- $q = 16$ traits, $m = 20,000$ markers
- Select 0 to 500K genotypes randomly (gen. 3)

Analysis

- Imputation of M_1
 - Factor A^{11} , minimise fill-in
 - Triangular solve: $A^{11} M_1 = -A^{12} M_2$
- Solving MME
 - PCG with diagonal pre-conditioner
- Parameterisation
 - MV16: standard multivariate
 - PCr: genetic principal components

Computing

- Intel Xeon CPU E5-2697, 28 cores @ 2.6Ghz, 512GB RAM
- FORTAN + OpenMP
- Multi-threaded Intel Math Kernel Library
 - BLAS, sparse BLAS, PARDISO
- In core storage
 - Dense: M_2 (4 to 75 GB), $M_1' A^{11} M_1$ (3 GB)
 - Sparse: CSR format

Stepwise product: Coefficient matrix \times vector

Need to calculate product for each PCG iterate:

- Partition according to type of effects

$$\begin{bmatrix} C_{bb} & C_{bu} & C_{b\alpha} \\ C'_{bu} & C_{uu} & C_{u\alpha} \\ C'_{b\alpha} & C'_{u\alpha} & C_{\alpha\alpha} \end{bmatrix} \begin{bmatrix} r_b \\ r_u \\ r_\alpha \end{bmatrix} = \begin{bmatrix} q_b \\ q_u \\ q_\alpha \end{bmatrix}$$

$$\begin{aligned} C_{\alpha\alpha} &= (I_q \otimes M_2') Z_2' R_2^{-1} Z_2 (I_q \otimes M_2) \\ &\quad + \Sigma_u^{-1} \otimes M_1' A^{11} M_1 + \Sigma_\alpha^{-1} \otimes D^{-1} \\ C_{u\alpha} &= \Sigma_u^{-1} \otimes A^{12} M_2 \\ C_{b\alpha} &= X_2' R_2^{-1} Z_2 (I_q \otimes M_2) \\ C_{uu} &= Z_1' R_1^{-1} Z_1 + \Sigma_u^{-1} \otimes A^{11} \\ C_{bu} &= X_1' R_1^{-1} Z_1 \\ C_{bb} &= X' R^{-1} X \end{aligned}$$

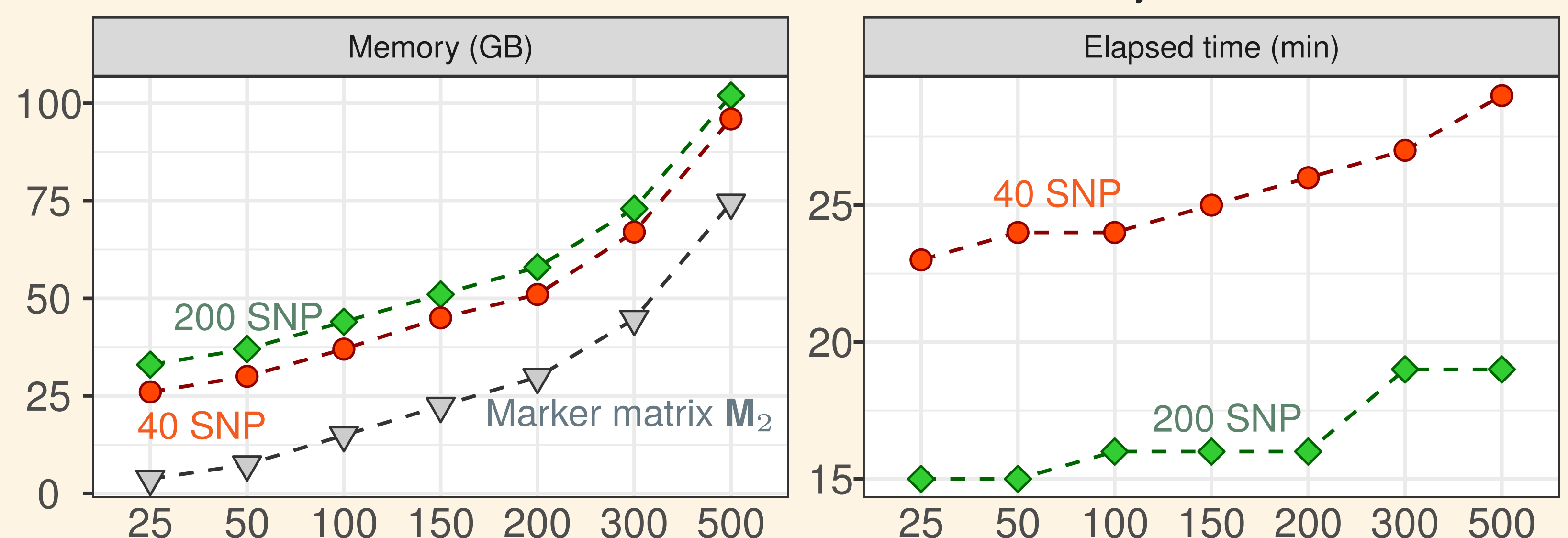
- Evaluate in steps – avoid forming dense matrices $C_{b\alpha}$, $C_{u\alpha}$ and $C_{\alpha\alpha}$

$$\begin{aligned} \begin{bmatrix} q_b \\ q_u \end{bmatrix} &:= \begin{bmatrix} C_{bb} & C_{bu} \\ C_{ub} & Z_1' R_1^{-1} Z_1 \end{bmatrix} \begin{bmatrix} r_b \\ r_u \end{bmatrix} & \left\{ \begin{aligned} q_u &:= q_u + (\Sigma_u^{-1} \otimes I_{n_1}) w \\ w &:= (I_q \otimes M_1' A^{11} M_1) r_\alpha \\ q_\alpha &:= (\Sigma_u^{-1} \otimes I_m) w \\ w &:= (I_q \otimes D^{-1}) r_\alpha \\ q_\alpha &:= q_\alpha + (\Sigma_\alpha^{-1} \otimes I_m) w \end{aligned} \right. & \left\{ \begin{aligned} t &:= Z_2' R_2^{-1} Z_2 t \\ t &:= t + Z_2' R_2^{-1} X_2 r_b \\ w &:= (I_q \otimes A^{21}) r_u \\ t &:= t + (\Sigma_u^{-1} \otimes I_{n_2}) w \\ q_\alpha &:= q_\alpha + (I_q \otimes M_2') t \end{aligned} \right. \\ t &:= (I_q \otimes M_2) r_\alpha \\ q_b &:= q_b + X_2' R_2^{-1} Z_2 t \\ w &:= (I_q \otimes A^{12}) t \\ w &:= w + (I_q \otimes A^{11}) r_u \end{aligned}$$

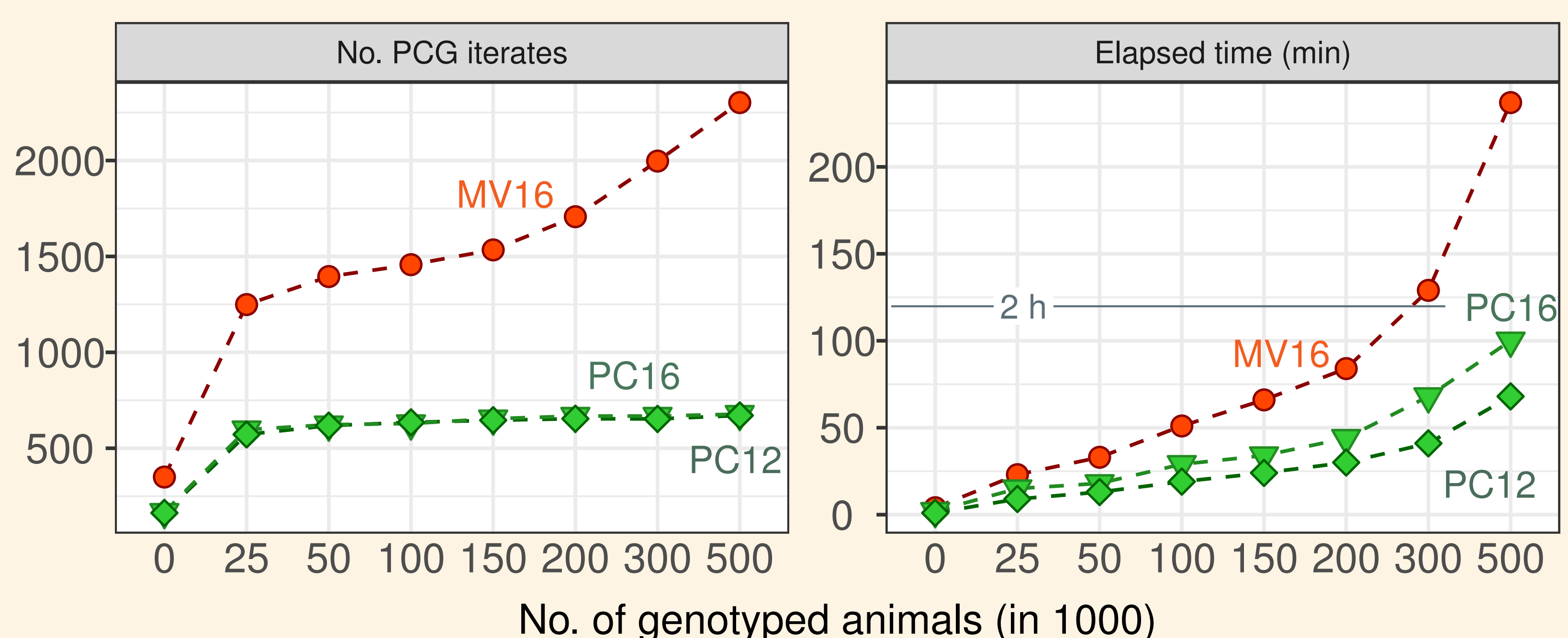
Results: Computational demands

Imputation of marker counts for non-genotyped animals

Process 40 or 200 SNP simultaneously



Iterative solution of mixed model equations



Conclusions

- Hybrid model readily implemented, good multi-thread performance
 - exploit highly optimized ‘off-the-shelf’ routines
- Trade-off between set-up & solution steps
 - HM best for large no.s of genotypes/small(ish) no.s of markers
 - \Rightarrow identify reduced marker set; fit residual polygenic effects?
- Parameterisation to genetic principal components is advantageous

