Bayesian calibration of earth systems models

Lev Tarasov[1], Radford Neal[2], and W. R. Peltier[2]

(1) Memorial University of Newfoundland, (2) University of Toronto, Canada

(1) Context: noisy data and model with significant computational cost

- Data: Relative Sea Level (RSL), geodetic (surface uplift), ice margin chronology, paleo-lake levels (strandlines),...
- Model: MUN/UofT Glacial Systems Model (GSM): 3D thermo-mechanically coupled ice-sheet model, visco-elastic bedrock response, surface drainage solver,...
- 32 ensemble parameters, non-linear system, large heterogeneous noisy constraint data set

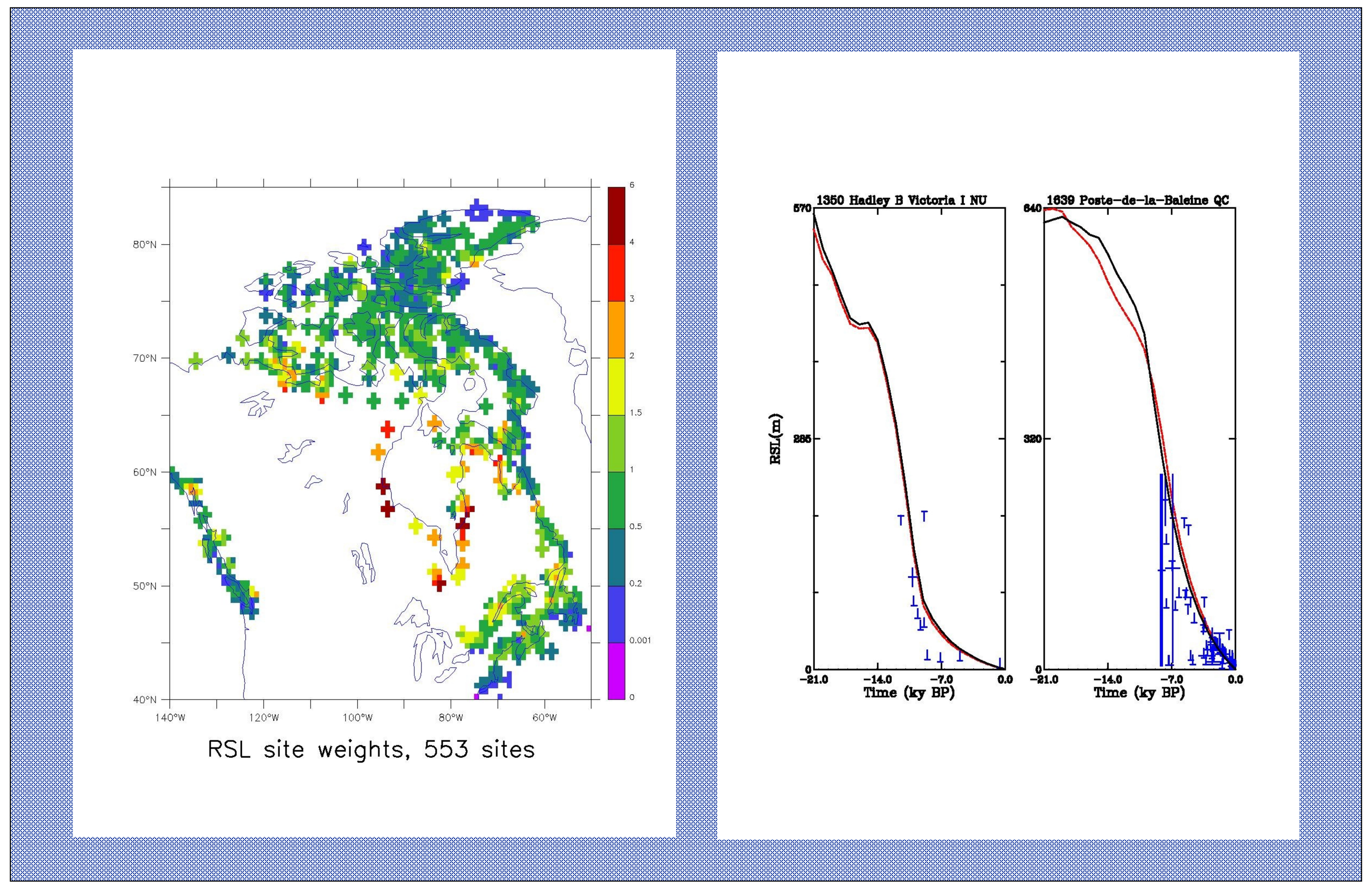
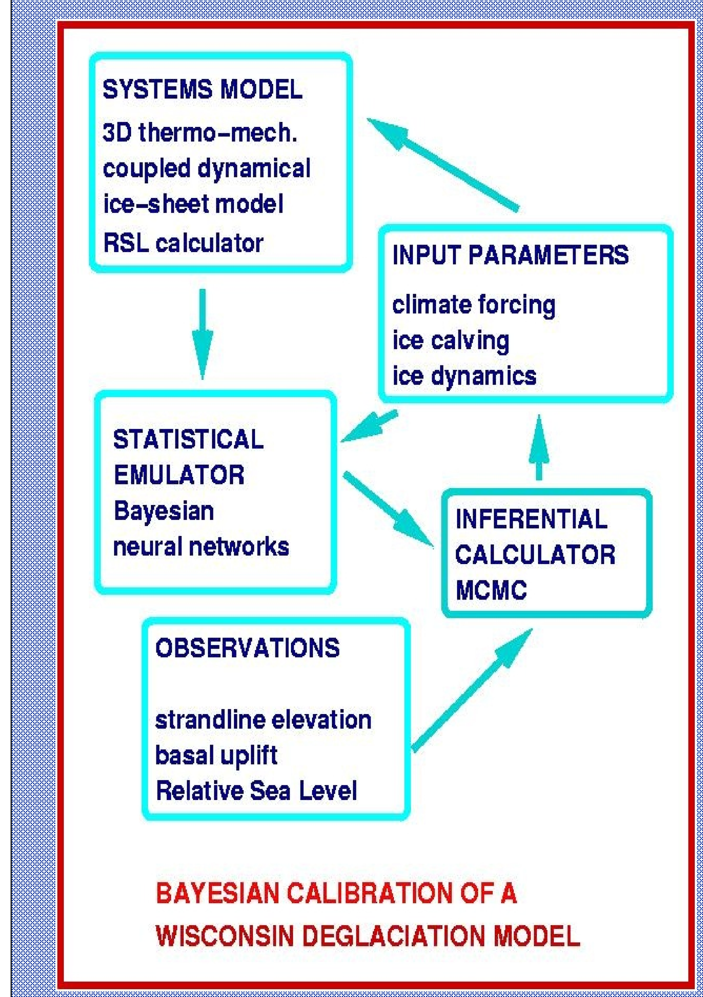


Figure 1. RSL site weights and example data

(2) Calibration procedure

Sample over posterior probability distribution for the ensemble parameters given fits to observational data using Markov Chain Monte Carlo (MCMC) methods



Calibration details

- The expectation of the posterior distribution of possible models (GSM) given constraint data set D is given by:

$$
\begin{aligned}
<GSM|D> &= \int GSM(y)P(y|D)dy \\
&= \int GSM(y)P(D|y)Z^{-1}P(y)dy \quad ,
\end{aligned}
$$

where P(y) is the prior probability distribution for the ensemble parameter vector y. The normalization constant Z is a function only of D and is therefore ignored.

- In MCMC sampling, this becomes

$$<GSM|D> \simeq N^{-1} \sum_{MCMC(P(y),P(D|y))} GSM(y) \quad ,$$

for N samples. However, given that $P(D|y)_{\text{networks}}$ employed in the MCMC sampling is not entirely accurate, and given that further observations (DML) are used to score the final results, use:

$$<GSM|D> \simeq N^{-1} \sum_{y(MCMC)} \{GSM(y) \cdot P(DML|y)_{\text{GSM}} P(D|y)_{\text{GSM}} / P(D|y)_{\text{networks}}\}$$

1

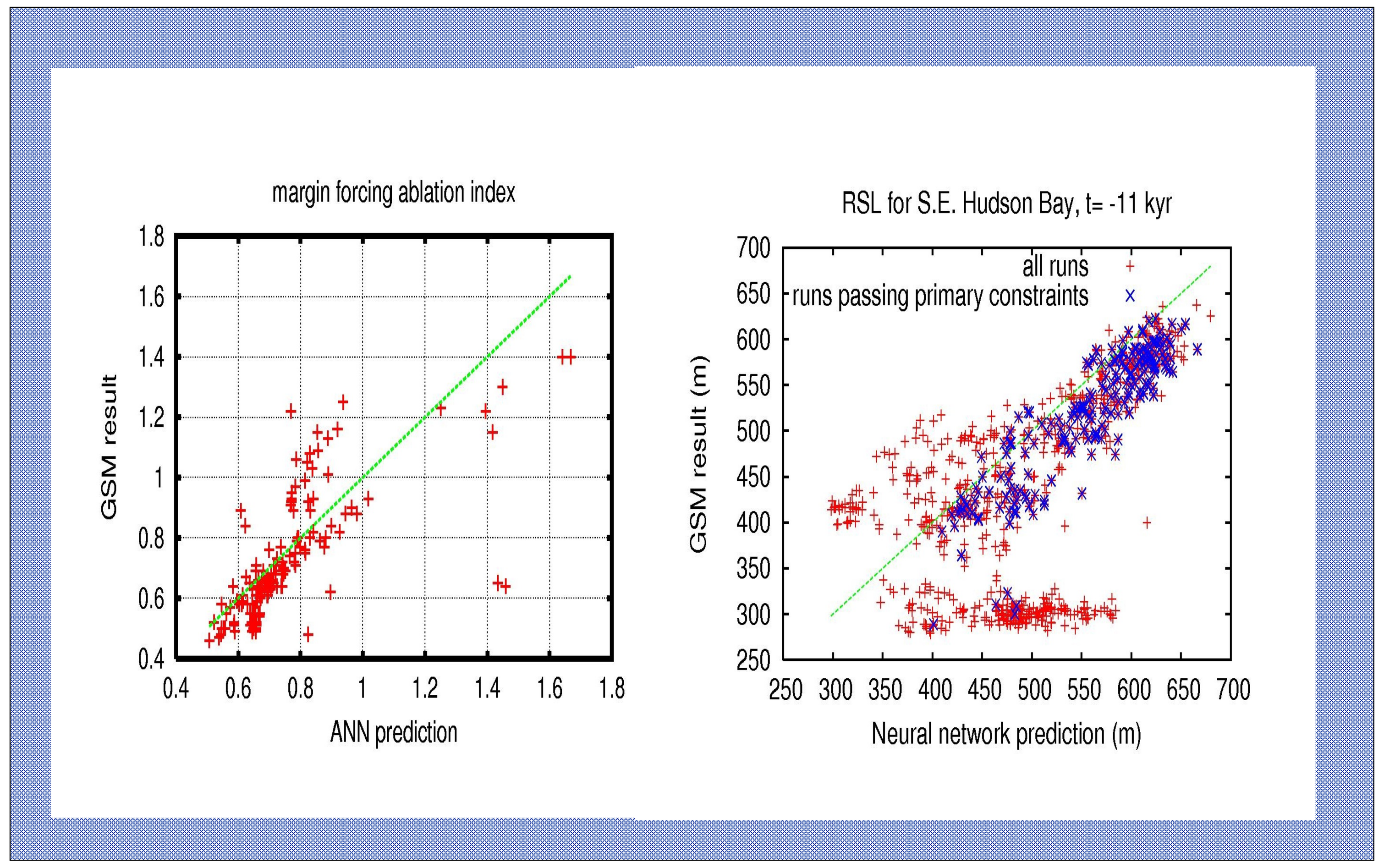


Figure 3a. Model results versus neural network predictions

(3) Calibration validation: neural networks

- Networks generally captured most of the model response
- RSL networks had the weakest fits due to large regional coverage and associated complexity of response
- Nevertheless, overall misfit prediction was reasonably accurate when RSL network was not overloaded

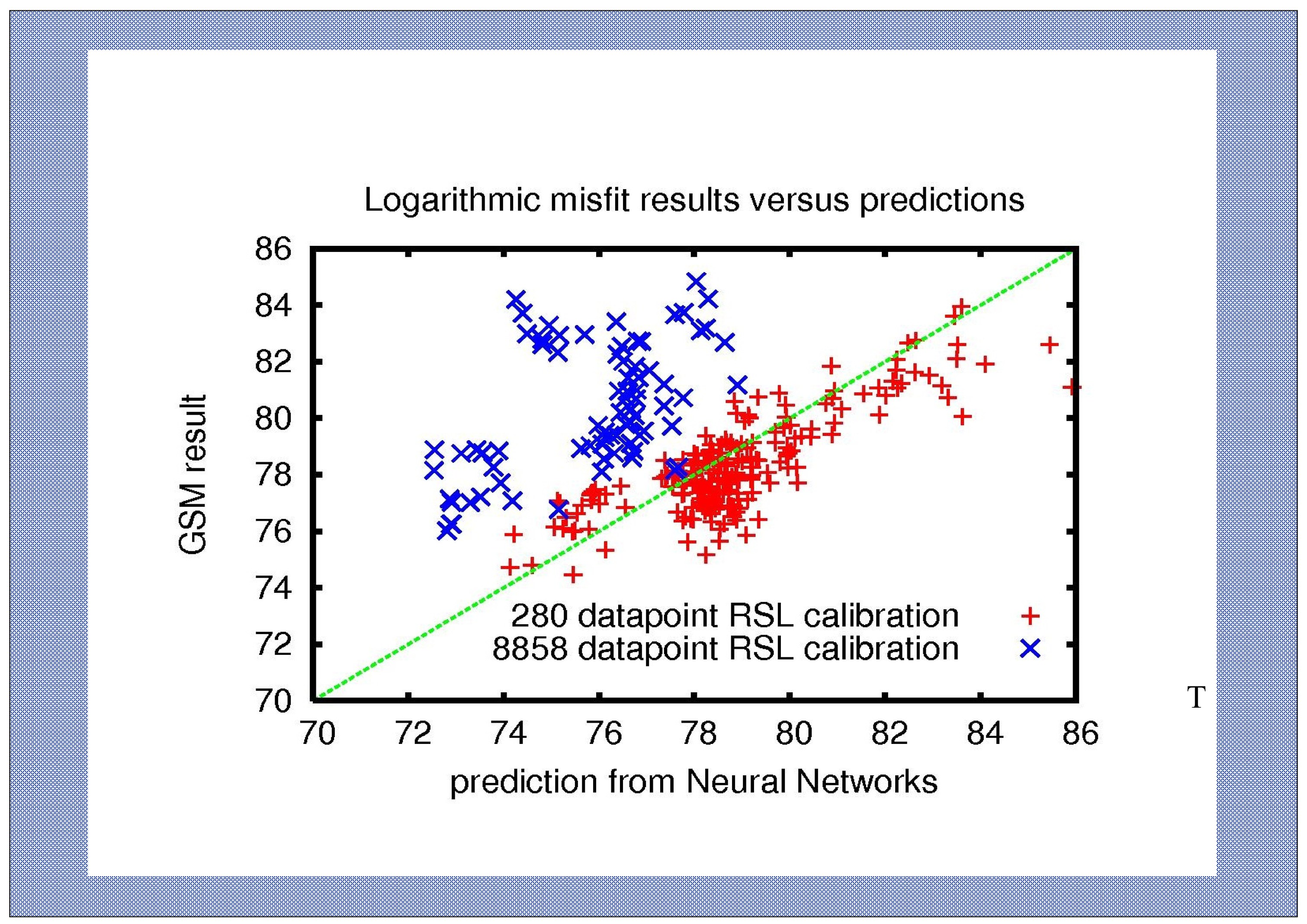


Figure 3b. Neg. scaled Logliklihood: model versus network

(4) Calibration Performance: MCMC

- MCMC chains sometimes get stuck around local minima
- Overall, MCMC sampling produced a much higher density of better fitting models than that of an ensemble with a Latin Hypercube set of parameters from the prior distribution (“random” in figure below).

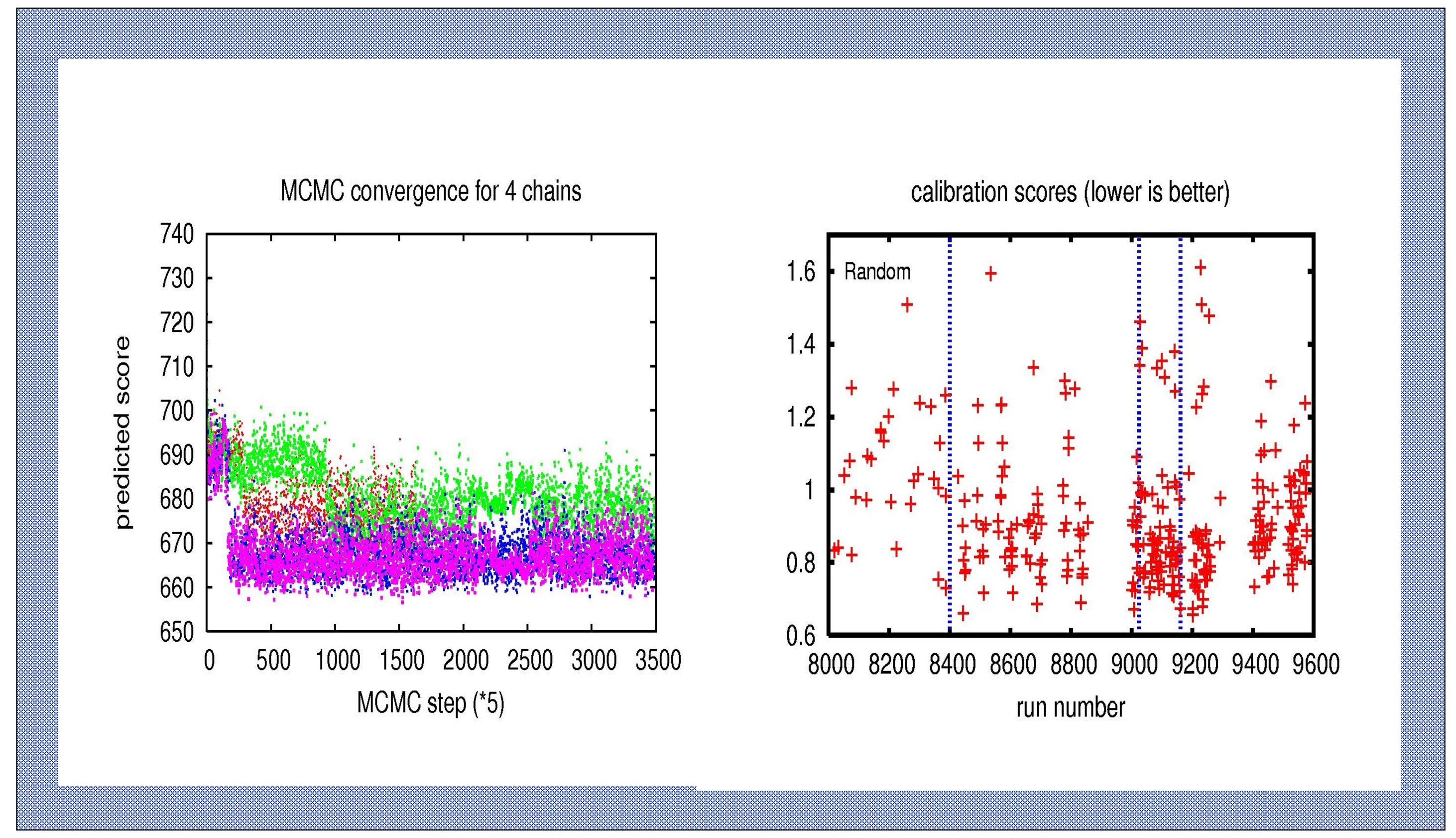


Figure 4. Full misfit metric values versus model runs

(5) Some lessons

- Start with kitchen sink -> shrink parameter set (using automatic relevance determination)
- Disaggregate poorly performing neural networks
- Start with a reduced constraint set and run multiple chains (10+). Consider filtering the constraint set.
- Issues: priors, error models for constraint data, aggregated metrics, and extra constraints (physicality and model stability)
- DATA+MODEL+CALIBRATION= MEANINGFUL PROBABILITY DISTRIBUTION FOR MODEL PREDICTIONS

References

Neal, R.M. (2003), Slice sampling, Ann. of Statis., 31, 705-767

Tarasov, L., and W. R. Peltier (2005), Arctic freshwater forcing of the Younger Dryas cold reversal, Nature, 435, 662-665