

1. Introduction and Aims

Understanding Extracellular Signal-Regulated Kinase (ERK) activation kinetics is crucial for understanding cellular processes such as proliferation and differentiation [2]. Ordinary Differential Equation (ODE) models of the Epidermal Growth Factor (EGF) stimulated ERK signalling pathway depend on over 50 unknown parameters, yet Vysheirsky et al. were recently able to use Bayesian inference to conclusively select one of four competing models based on timecourse data of ERK alone [2]. Additionally, Gutenkunst et al. have done important work demonstrating that **most of the degrees-of-freedom in ODE systems biology models are essentially unimportant** [1]. They term **this apparently universal property of biochemical systems models 'sloppiness'**. In this work we find sloppiness in two ERK signalling models, and analyse this in the context of parameter inference.

2. Methods

Following [1], we define a sum-of-squared-residuals log-likelihood, and **analyse the sloppiness with a local perturbation analysis**. To minimise errors due to a lack of data, we simulate 4000 timepoints per species, adding a Gaussian noise term.

$$\log p(x|\theta) = -\frac{1}{2N_s N_t} \sum_{s,t} \left(\frac{y_{s,t} - x_{s,t}(\theta)}{\sigma_s} \right)^2$$

Model outputs

$$\frac{dx}{dt} = f(x, \theta, t)$$

Simulated data

Figure 1: Phosphorylated ERK expression profiles plotted over 4000 timepoints for each ERK signalling model. Similar profiles were generated for every species

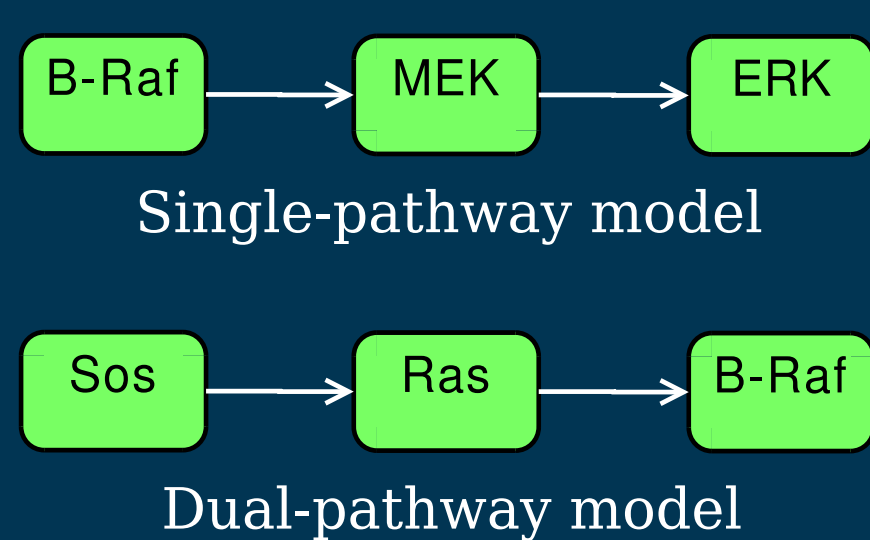
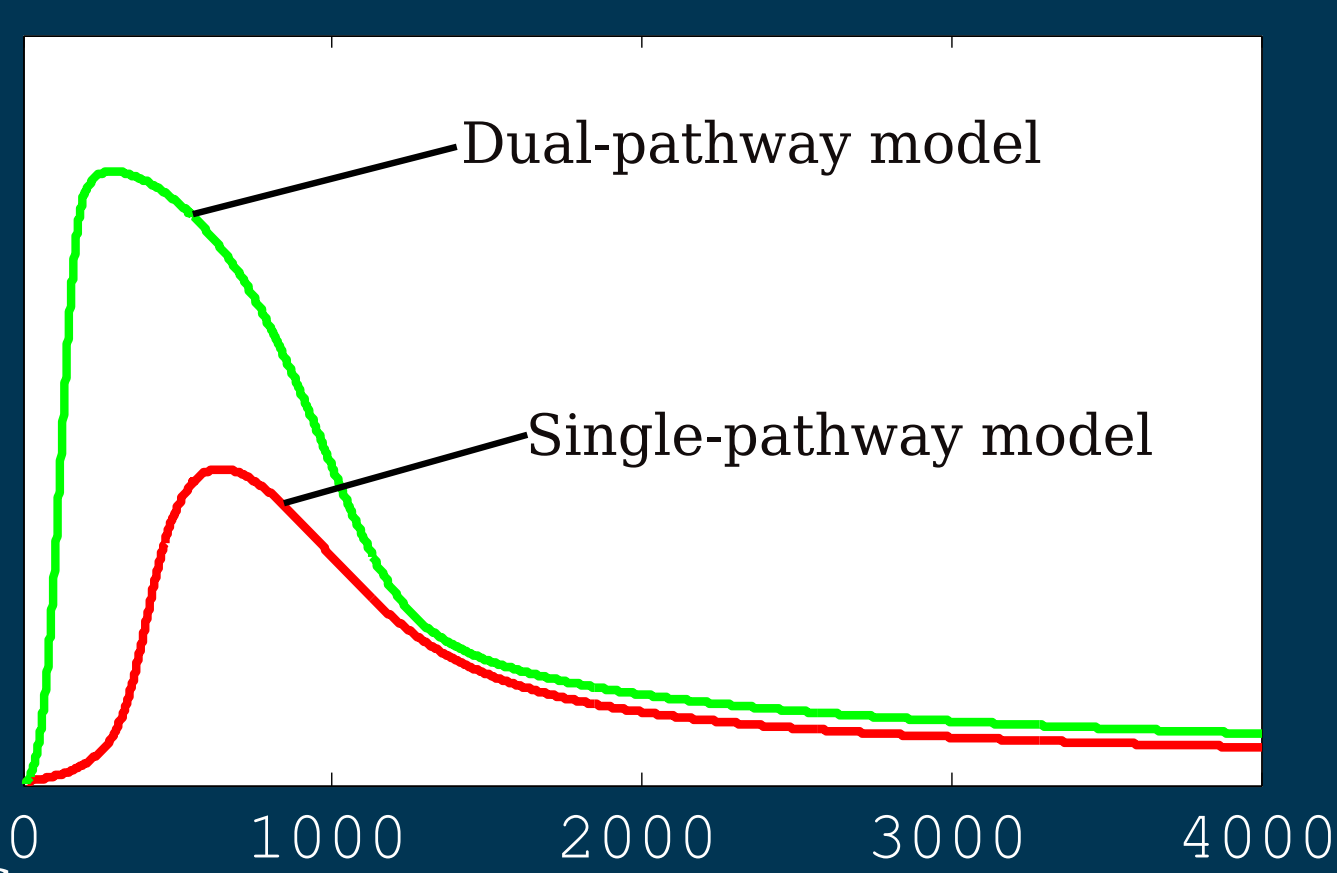


Figure 3: Species whose activation and deactivation parameters are large components of each model's respective main eigenvector.

3. Sloppiness

The eigenvalue spectra of the Fisher Information matrices for the fully observed models are **sloppy**. The leading eigenvalue in each case is an order of magnitude larger than the next largest, and the **magnitudes of**

all the eigenvalues range over several decades in log-space. In each case, the 6 smallest eigenvalues in the models represent parameters controlling PKA and EPAC, which were set to zero in our simulations.

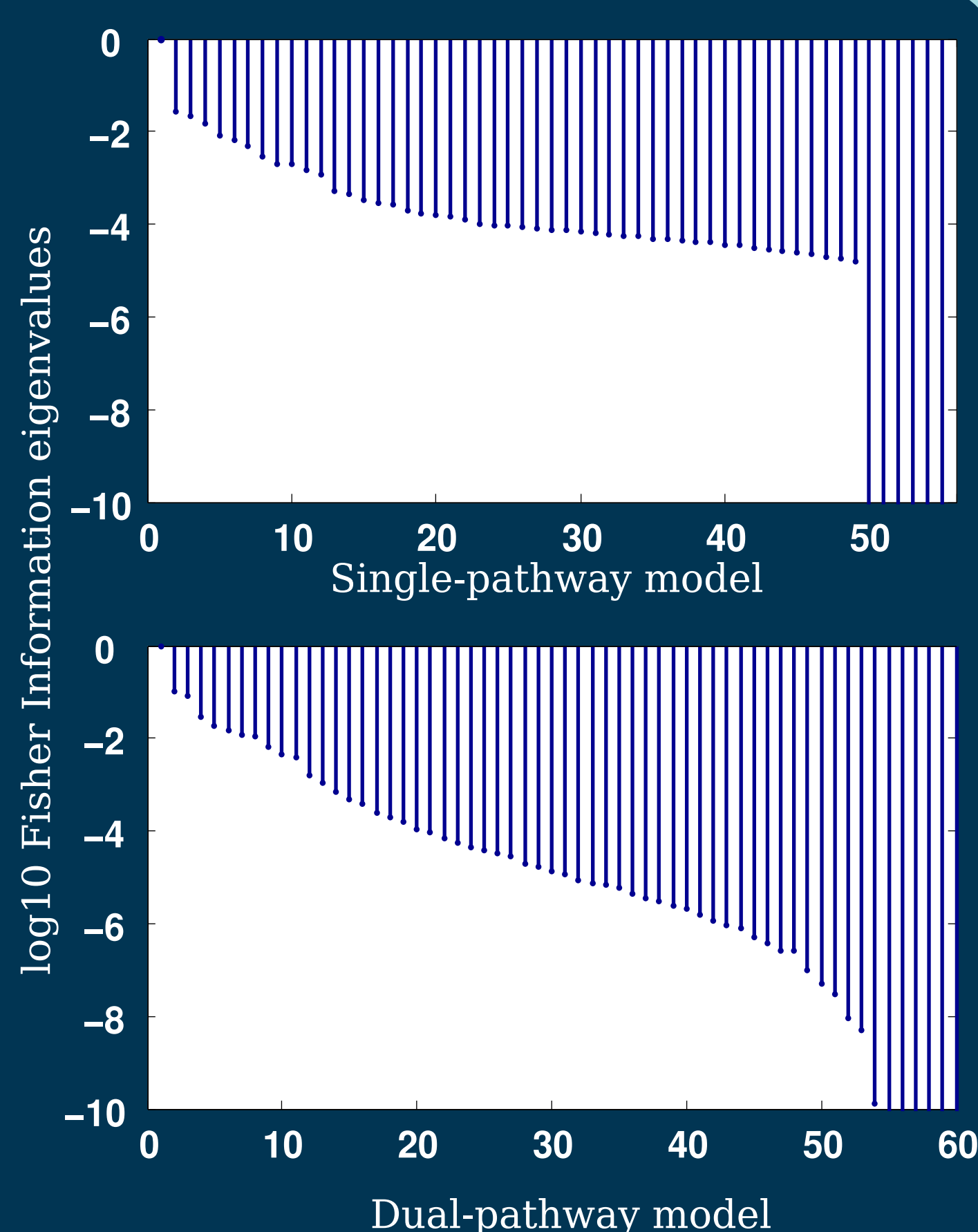


Figure 2: The sloppy spectra of the log-eigenvalues obtained from a decomposition of the Fisher Information matrices for the fully observed models.

$$\frac{\partial^2 \log p(x|\theta)}{\partial \log \theta_i \partial \log \theta_j} \Big|_{\theta^*} = \frac{1}{N_s N_t} \sum_{s,t} \frac{1}{\sigma_s} \frac{\partial x_{s,t}}{\partial \log \theta_i} \frac{1}{\sigma_s} \frac{\partial x_{s,t}}{\partial \log \theta_j} \Big|_{\theta^*} = J_{ij}$$

Equation 2: The second derivative of the log-likelihood, evaluated at our operating point, is equivalent to the Fisher Information matrix [3]. Performing a singular value decomposition of this matrix, we obtain the sloppy spectra of eigenvalues depicted in the figures on the left.

4. Fisher Information

Thanks to the Cramer-Rao bound, we can interpret the trace of the inverse of the Fisher Information matrix as a lower bound on the parameter variance estimates [3]. We use the inverse of the largest eigenvalue as a crude measure of this bound, and therefore we **interpret the largest eigenvalue as a measure of the Fisher Information content of a set of observations**. To gain an insight into how this measure varies across the models, we repeat our analysis for single species observations, for every species in the model. The results, depicted below, show a **striking variation in information content of species across the models**.

Equation 3: The Cramer-Rao lower bound on the variance of our estimate of the parameters

$$\left\langle [\hat{\theta} - \theta]^T [\hat{\theta} - \theta] \right\rangle \geq \text{Tr} [J^{-1}(\theta)]$$

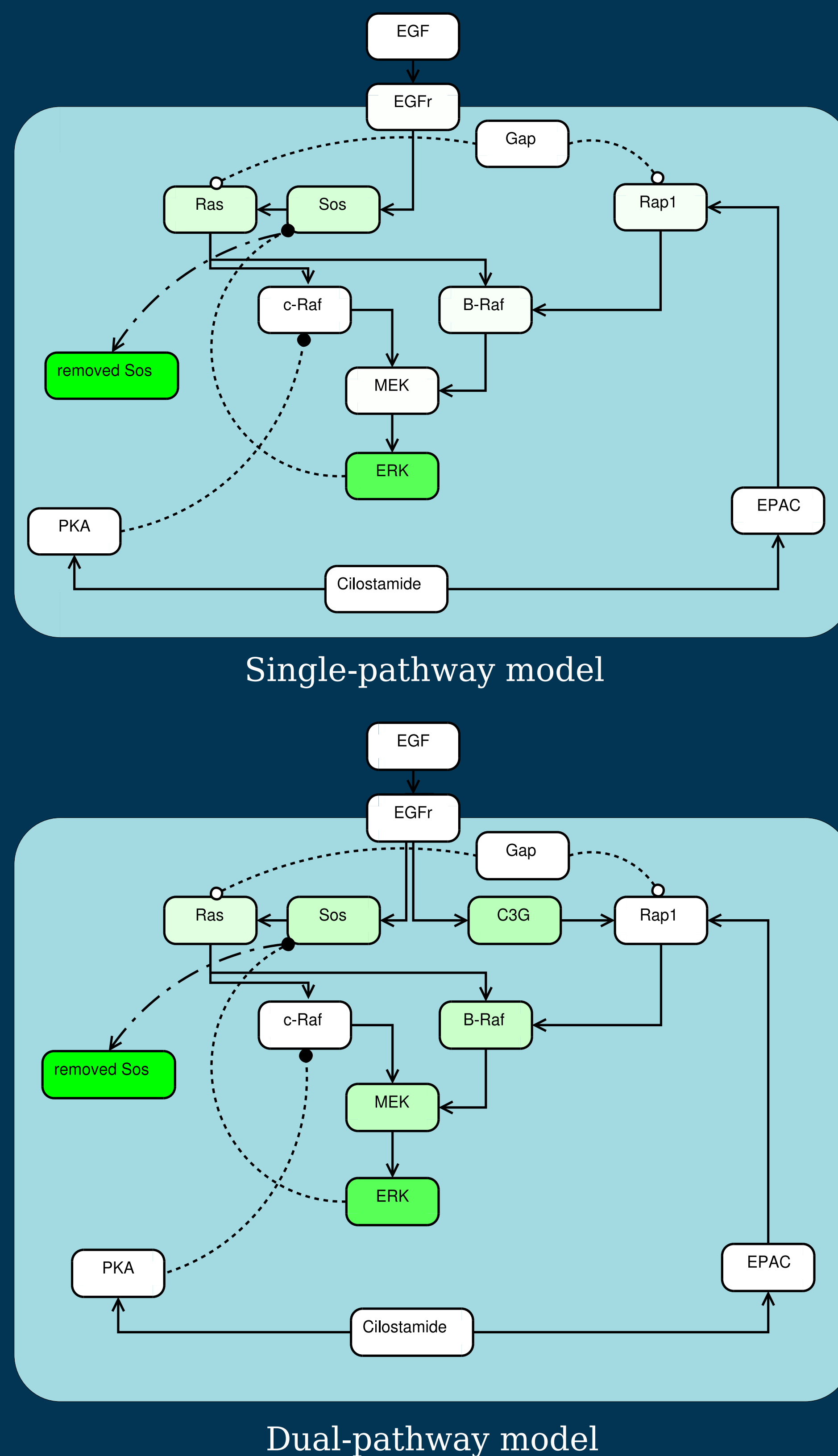


Figure 4. Depiction of the magnitude of the leading eigenvalue of the Fisher Information matrix for single-species observation of the models. Only the activated forms of the proteins are considered here - in each case, the inactivated form was less informative. The species are coloured green, with intensity proportional to the size of the eigenvalue. In both models, removed Sos is the most informative species, with ERK a close second. C-Raf is uninformative due to strong Cilostamide-activated PKA inhibition.

5. Conclusions

Both of the ERK signalling models considered in this work are **sloppy**, providing further support to Gutenkunst's contention that sloppiness is a universal feature of biochemical systems models. We find qualitatively similar sloppy eigenspectra to those found in [1], and we use the **parameters implicated by the main eigenvectors to identify potentially important subnetworks**. Finally, we find that **some single species observations have notably larger Fisher Information than others**. Further work is needed to investigate the implications of this analysis for model identifiability.

References

- Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., and Sethna, J. P. PLoS Comput Biol 3(10), e189 Oct (2007).
- Vysheirsky, V., Tian-Rui, X., Girolami, M., Gormand, A., Baillie, S., Houslay, M., and Kolch, W. Technical report, University of Glasgow, Department of Computing Science, (2008).
- Scharf, L. and McWhorter, L. Statistical Signal and Array Processing, 1992. Conference Proceedings., IEEE Sixth SP Workshop on , 5-8 Oct (1992).

Acknowledgments

This work was supervised by Mark Girolami, who is funded by an EPSRC Advanced Research Fellowship. The author is part of the Doctoral Training Centre in Proteomic Technologies, funded by EPSRC, BBSRC and MRC.