Exploring experimental designs for network inference using perturbations and a Bayesian sequential learning strategy

Christopher Penfold\(^1\) and David Wild and PRESTA

\(^1\)C.A.Penfold@warwick.ac.uk

April 1, 2009
Table of contents

1. Introduction
   - Who/What/Why/How?

2. Simple Networks
   - Realistic in Silico Networks

3. Results
   - Sequential Learning and Active Intervention
Who/What?

Who?

PRESTA (Plant Response to Environmental Stress in Arabidopsis)
Who/What?

Who?

PRESTA (Plant Response to Environmental Stress in Arabidopsis)

- Theorists and Biologists.
Who? PRESTA (Plant Response to Environmental Stress in Arabidopsis)
  - Theorists and Biologists.
  - Warwick, Essex and Exeter.

What?
Who?

PRESTA (Plant Response to Environmental Stress in Arabidopsis)

- Theorists and Biologists.
- Warwick, Essex and Exeter.

What?

Stress Response Network
Who? PRESTA (Plant Response to Environmental Stress in Arabidopsis)
- Theorists and Biologists.
- Warwick, Essex and Exeter.

What? Stress Response Network
- Multiple Datasets
Who/What?

Who?

PRESTA (Plant Response to Environmental Stress in Arabidopsis)
- Theorists and Biologists.
- Warwick, Essex and Exeter.

What?

Stress Response Network
- Multiple Datasets
- Prior Knowledge
**Who?**

1. PRESTA (**P**lant **R**esponse to **E**nvironmental **S**Tress in **A**rabidopsis)
   - Theorists and Biologists.
   - Warwick, Essex and Exeter.

**What?**

1. Stress Response Network
   - Multiple Datasets
   - Prior Knowledge
   - Large Datasets
Christopher Penfold and David Wild and PRESTA

Exploring experimental designs for network inference using perturbations.
Why?

1. Targets for improving stress response
Why?

1. Targets for improving stress response
2. (Help) design next experiments
Why?

1. Targets for improving stress response
2. (Help) design next experiments
3. Example of multidisciplinary approach
How: Bayesian State Space Models

How: Bayesian State Space Models

Figure: Graphical Model for the BSSM\textsuperscript{a}

\textsuperscript{a}Beal et al., Bioinformatics 21:349-356, 2005.
How: Bayesian State Space Models

Figure: Graphical Model for the BSSM\textsuperscript{a}

\[ x_t = Ax_{t-1} + By_{t-1} + w_t, \]
\[ y_t = Cx_t + Dy_{t-1} + v_t. \]

\textsuperscript{a}Beal et al., Bioinformatics 21:349-356, 2005.
How: Bayesian State Space Models

Figure: Graphical Model for the BSSM\textsuperscript{a}

\[ x_t = A x_{t-1} + B y_{t-1} + w_t, \]
\[ y_t = C x_t + D y_{t-1} + v_t. \]

\[ y_t = (CB + D)y_{t-1} + r_t. \]

\textsuperscript{a}Beal et al., Bioinformatics 21:349-356, 2005.
Figure: ODE model of Zak et al.\textsuperscript{1}

Active Interventions

1. Knockouts ✓
Active Interventions

1. Knockouts ✓
2. Over expression ✓
Active Interventions

1. Knockouts ✓
2. Over expression ✓
3. Gene Duplication ✓
Active Interventions

1. Knockouts ✓
2. Over expression ✓
3. Gene Duplication ✓
4. Silencing ✓
Active Interventions

1. Knockouts ✓
2. Over expression ✓
3. Gene Duplication ✓
4. Silencing ✓
5. Additional links

Christopher Penfold and David Wild and PRESTA
Exploring experimental designs for network inference using pert
Expression Profiles: Knockouts

Figure: **Knockout A**: downstream profiles B & C
Expression Profiles: Knockouts

Figure: **Knockout A**: downstream profiles B & C

Figure: **Knockout C**: downstream profiles G & J
Expression Profiles: Overexpressors

Figure: Overexpress C: downstream genes G & J

Figure: Overexpress F: downstream genes B & D
Performance

- ROC curve
  - True Positives
  - False Positives
Performance

- ROC curve
  - True Positives
  - False Positives
Performance

1. ROC curve
   - True Positives
   - False Positives

2. Area Under ROC Curve (AUC)
Sequential Learning

1. Train Wild Type
2. Train Mutant Systems using WT as Prior
Sequential Learning

1. Train Wild Type
2. Train Mutant Systems using WT as Prior

![Diagram]

Christopher Penfold and David Wild and PRESTA

Exploring experimental designs for network inference using perturbations.
Sequential Learning

1. Train Wild Type
2. Train Mutant Systems using WT as Prior

- Knockout (Dark Blue)
- Overexpressor (Light Blue)
- Gene Duplication (Green)
- Silencing (Orange)
- WT (Red)
## Sequential Learning vs. Active Intervention

![Graph showing comparison between Sequential Learning and Active Intervention](image_url)

Christopher Penfold and David Wild and PRESTA

*Exploring experimental designs for network inference using perturbations*
Sequential Learning vs. Active Intervention

Christopher Penfold and David Wild and PRESTA
Exploring experimental designs for network inference using perturbations
Using Prior **CAN** increase AUC

- Using information rich priors *does not* necessarily allow us to recapture all of the information gleaned from more time points.
Summary

1. Using Prior **CAN** increase AUC
   - Using information rich priors *does not* necessarily allow us to recapture all of the information gleaned from more time points

2. Active intervention alone **CAN** increase AUC
Summary

1. Using Prior **CAN** increase AUC
   - Using information rich priors *does not* necessarily allow us to recapture all of the information gleaned from more time points

2. Active intervention alone **CAN** increase AUC
   - Knockouts *not necessarily* informative
Summary

1. Using Prior **CAN** increase AUC
   - Using information rich priors *does not* necessarily allow us to recapture all of the information gleaned from more time points

2. Active intervention alone **CAN** increase AUC
   - Knockouts **not necessarily** informative
   - Overexpression **not necessarily** informative
Summary

1. Using Prior CAN increase AUC
   - Using information rich priors does not necessarily allow us to recapture all of the information gleaned from more time points
2. Active intervention alone CAN increase AUC
   - Knockouts not necessarily informative
   - Overexpression not necessarily informative
3. Actual data: VBSSM picks out “hubs” (increases phenotype discovery from \( \sim 1\% \) to \( \sim 40\% \)).
Future Work

\[ M_1 \rightarrow M_2 \rightarrow M_3 \]

\[ \beta_1 \rightarrow D_1 \rightarrow \beta_2 \rightarrow D_2 \rightarrow \beta_3 \rightarrow D_3 \]

Future Work

Learn Jointly Over Datasets a la Werhli and Husmeier\textsuperscript{a}

Future Work

1. Learn Jointly Over Datasets *a la* Werhli and Husmeier$^a$
   - MCMC sampling method for BSSM.

---

Future Work

1. Learn Jointly Over Datasets \textit{a la} Werhli and Husmeier\textsuperscript{a}
   - MCMC sampling method for BSSM.

2. Gaussian Process Latent Variable Model \textit{a la} Klemm and Ghahramani

PRESTA
David Wild
Emma Cooke
Stuart McHattie
Sapna Sharma