# Structural and practical identifiability of ERK kinetics

(joint work with Lewis Marsh, Helen Byrne and Heather Harrington)

Emilie Dufresne, University of York

DART XI, London, June 2023

# The model and experimental set-up 1

## The Full ERK Model

dual phosphorylation of ERK by MEK

intermediate complexes

$$
\begin{tikzcd}
E+S_0 \arrow[r, shift left, "k_{f_1}"] & C_1 \arrow[l, shift left, "k_{r_1}"] \arrow[r, "k_{c_1}"] & C_2 \arrow[r, "k_{c_2}"] \arrow[d, shift right, "k_{r_2}"'] & E+S_2 \\
& & E+S_1 \arrow[u, shift right, "k_{f_2}"'] &
\end{tikzcd}
$$

Cell division,
Cell specialisation,
Cell death

parameter space: $\Theta = \mathbb{R}^6_{>0}$

$\ni$

$(k_{f_1}, k_{r_1}, k_{c_1}, k_{c_2}, k_{f_2}, k_{c_2})$

MEK variants: Wild type, E203K, F53S, Y130, SSDD

E

E203K: cancer

F53S, Y130: developmental abnormalities

SSDD: often used as approximation of wild type.

Initial conditions: $C_1(t=0) = C_2(t=0) = S_1(t=0) = S_2(t=0) = 0$

$S_0(t=0) = S_{tot} = 5\,\mu M$, $E(t=0) = E_{tot} = 0.65\,\mu M$

# The model and experimental set-up 2

$$\frac{dS_0}{dt} = -k_{f_1} E \cdot S_0 + k_{r_1} C_1$$

$$\frac{dC_1}{dt} = k_{f_1} E \cdot S_0 - (k_{r_1} + k_{c_1}) C_1$$

$$\frac{dC_2}{dt} = k_{c_1} C_1 - (k_{r_2} + k_{c_2}) C_2 + k_{f_2} E \cdot S_1$$

$$\frac{dS_1}{dt} = -k_{f_2} E \cdot S_1 + k_{r_2} C_2$$

$$\frac{dS_2}{dt} = k_{c_2} C_2$$

$$\frac{dE}{dt} = -k_{f_1} E \cdot S_0 + k_{r_1} C_1 - k_{f_2} E \cdot S_1 + (k_{r_2} + k_{c_2}) C_2$$

$$\begin{tikzcd}
E+S_0 \arrow[r, shift left, "k_{f_1}"] & C_1 \arrow[l, shift left, "k_{r_1}"] \arrow[r, "k_{c_1}"] & C_2 \arrow[r, "k_{c_2}"] \arrow[d, shift right, "k_{r_2}"'] & E+S_2 \\
 & & E+S_1 \arrow[u, shift right, "k_{f_2}"'] &
\end{tikzcd}$$

**parameter space:** $\Theta = \mathbb{R}^6_{>0}$

**MEK variants:** Wild type, E203K, F53S, Y130, SSDD

E203K: cancer
F53S, Y130: developmental abnormalities
SSDD: often used as approximation of wild type.

**Conserved quantities:**

$S_{tot} = S_0 + S_1 + S_2 + C_1 + C_2$

$E_{tot} = E + C_1 + C_2$

**Time points** (in minutes) $\{1, 2, 3.25, 5, 10, 20, 40\}$ SSDD

$\{0.5, 2, 3.25, 3.75, 5, 10, 20\}$ others

**# of replicates:** $r=11$ wild type, $r=6$ SSDD, $r=5$ others

**Initial conditions:** $C_1(t=0) = C_2(t=0) = S_1(t=0) = S_2(t=0) = 0$

$S_0(t=0) = S_{tot} = 5\,\mu M$, $E(t=0) = E_{tot} = 0.65\,\mu M$

# Quasi-Steady-State-Approximation

**Basic idea:** At some timescales, the rate of change of some variables is negligible.

**Start:** heuristic arguments based on fast and slow reactions with some assumptions
→ Henri, Michaelis-Menten, Briggs and Haldane
1910's 1920's

**Making it rigourous:** Singular-Perturbation approach
↳ Heineken et al, then Segel and Slemrod
1960's
Shauer and Heinrich
1980's

**Algebraic approach:** Goeke and Walcher (2014) and then Goeke et al:
→ explicit formula for QSSA model and characterisation of parameters for which the approximation is accurate.

# Algebraic QSSA for the Full ERK Model

## The Rational ERK model: (using $E = E_{tot} - C_1 - C_2$)

$$\frac{dS_0}{dt} = \frac{-k_1 S_0}{\gamma_1 S_0 + \gamma_2 S_1 + 1}$$

$$\frac{dS_1}{dt} = \frac{-k_2 S_1 + (1-\pi)k_1 S_0}{\gamma_1 S_0 + \gamma_2 S_1 + 1}$$

$$\frac{dS_2}{dt} = \frac{\pi k_1 S_0 + k_2 S_1}{\gamma_1 S_0 + \gamma_2 S_1 + 1}$$

**parameter space:** $\Theta = \mathbb{R}^4_{>0} \times (0,1)$

$$\boxed{k_i = E_{tot} \frac{k_{f_i} k_{c_i}}{k_{c_i} + k_{r_i}} \qquad \pi = \frac{k_{c_2}}{k_{c_2} + k_{r_2}} \qquad \gamma_i = k_{f_i} \frac{k_{c_1} + k_{c_2}}{(k_{c_1} + k_{r_1})(k_{c_2} + k_{r_2})}}$$

## The Linear ERK Model: (using $E = E_{tot} - S_{tot} + S_0 + S_1 + S_2$)

$$\frac{dS_0}{dt} = -k_1 S_0$$

$$\frac{dS_1}{dt} = -k_2 S_1 + (1-\pi)k_1 S_0$$

$$\frac{dS_2}{dt} = \pi k_1 S_0 + k_2 S_1$$

**parameter space:** $\Theta = \mathbb{R}^2_{>0} \times (0,1)$

$\rightsquigarrow$ cannot be obtained via the singular perturbation approach

$\rightsquigarrow$ was in the Yeung et al. paper

# Structural identifiability

The model prediction map:

$$\varphi_{t_1,\dots,t_7,r} : \Theta \longrightarrow \mathbb{R}^{21r}$$

$$\theta \longmapsto \text{measurements of the 3 species } S_0, S_1, S_2 \text{ at 7 time points over } r \text{ replicates}$$

This induces an equivalence relation $\sim_{t_1,\dots,t_7,r}$ on $\Theta$:

$$\theta \sim_{t_1,\dots,t_7,r} \theta' \iff \varphi_{t_1,\dots,t_7,r}(\theta) = \varphi_{t_1,\dots,t_7,r}(\theta')$$

<u>Definition</u> The model is

globally identifiable $\iff$ for all $\theta \in \Theta$ $\quad |[\theta]_{\sim_{t_1,\dots,t_7,r}}| = 1$

generically identifiable $\iff$ for almost all $\theta \in \Theta$ $\quad |[\theta]_{\sim_{t_1,\dots,t_7,r}}| = 1$

locally identifiable $\iff$ for almost all $\theta \in \Theta$ $\quad |[\theta]_{\sim_{t_1,\dots,t_7,r}}| < \infty$

generically non-identifiable $\iff$ for almost all $\theta \in \Theta$ $\quad |[\theta]_{\sim_{t_1,\dots,t_7,r}}| = \infty$

# Structural identifiability of our 3 ERK models

→ Existing methods assume knowledge of the full trajetories

↳ a result of Sontag implies that $2m+1$ generic time points provide the same information

$m$ = dim. of $\Theta$

How do we know we have "good" time points?

**Linear ERK Model:** using the analytic solutions we can write the model prediction map explicitly and show it's injective for any choice of 3 time points

⇝ globally identifiabe

**Full and Rational ERK Model:** using SIAN we show that they are generalically identifiable. We also show that our initial conditions are generic.

↳ we may not have enough time points

# Structural identifiability via differential algebra

(Hong et al 2019)

SIAN and other methods based on differential algebra (like DAISY) rely on the differential Nullstellensatz:

(Bellu et al. 2007)

Ritt 1950, Seidenberg 1952

$\mathbb{K}$ differentially closed field.

$$\left\{\begin{array}{c}\text{diffentially closed subsets} \\ \text{of } \mathbb{K}^n\end{array}\right\} \xleftrightarrow{\text{1-1}} \left\{\begin{array}{l}\text{radical differential} \\ \text{ideal in a} \\ \text{differential ring}\end{array}\right\}$$

"abstract" solutions

1-1 ↕

"analytic solutions"

$$\left\{\begin{array}{l}\text{n-tuples of "analytic functions"} \\ \text{satisfying the system of differential equations}\end{array}\right\}$$

# Structural identifiability via differential algebra 2

For an ODE system of the form $\dot{x} = f(\theta, x)$ $x(0) = x_0$

$y = g(x)$

The differential ring is $\mathbb{C}(\theta)\{x, y\}$

polynomial ring in $x, y$ and their derivatives with coefficients in $\mathbb{C}(\theta)$ (rational functions of $\theta$)

The differential ideal is

$$I_\Sigma = \left\langle (\dot{x}_i - f_i)^{(j)}, (y_k - g_k)^{(j)} \;\middle|\; 1 \le i \le n,\ 1 \le k \le m,\ j \ge 0 \right\rangle$$

$I_\Sigma$ is prime, so $\mathbb{C}(\theta)\{x, y\}/I_\Sigma$ is a domain

# Structural identifiability via differential algebra 3

SIAN relies on the following key mathematical result:

$k$ is subfield generated by $\mathbb{C}\{y\} + I_\Sigma$

<u>Proposition</u> (Hong et al 2020)

- If $k = k(\theta)$ then the model is generically identifiable
- If the field extension $k \subseteq k(\theta)$ is algebraic then the model is locally identifiable.

$\rightsquigarrow$ Hong et al make this effective, and then present a probabilistic version <u>which</u> can then be used to study "realistic models". This is implemented in Maple.

# Practical Identifiability

Assumptions:
- the model is generically identifiable
- $\Psi(\theta, z)$ is the probability distribution of data with true parameter $\theta$
  $\rightsquigarrow$ for generic $z^*$, the MLE $\hat{\theta}(z^*)$ exists and is unique
- $d_\Theta$: reference metric on parameter space $\Theta$

$$U_\delta(z^*) := \{\theta \in \Theta \mid -\log \Psi(\theta, z^*) < \delta\}$$ ← $\delta$-confidence region

**Definition:** The model is practically identifiable for a data point $z^*$ at significance level $\alpha$ iff the confidence region $U_\delta(z^*)$ is bounded with respect to $d_\Theta$, where $\delta = -\log \Psi(\hat{\theta}(z^*), z^*) - \log k^*$ and

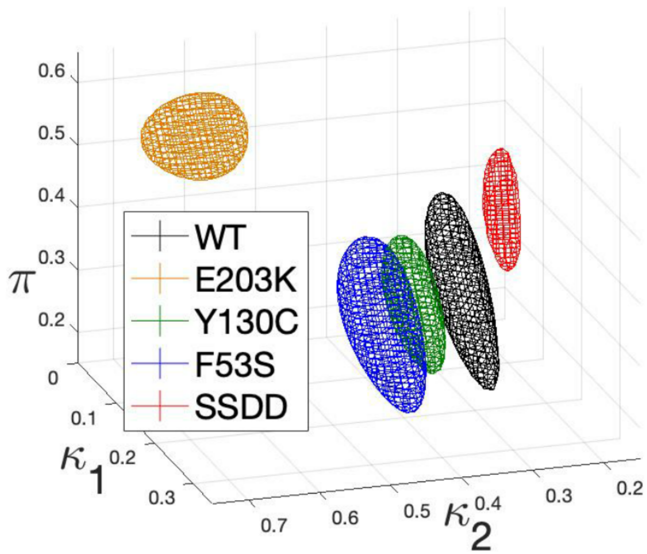
$$\alpha = \Pr\left( \frac{\Psi(\hat{\theta}(z^*), \hat{z})}{\max_{\theta \in \Theta} \Psi(\theta, \hat{z})} < k^* \;\middle|\; \hat{z} \text{ is data with true parameter } \hat{\theta}(z^*) \right)$$

# Practical identifiability of ERK kinetics

→ $d_{\Theta}$ is the Euclidean metric

→ measurement error is additive Gaussian noise with covariance a multiple of identity and mean the image of the model prediction map.

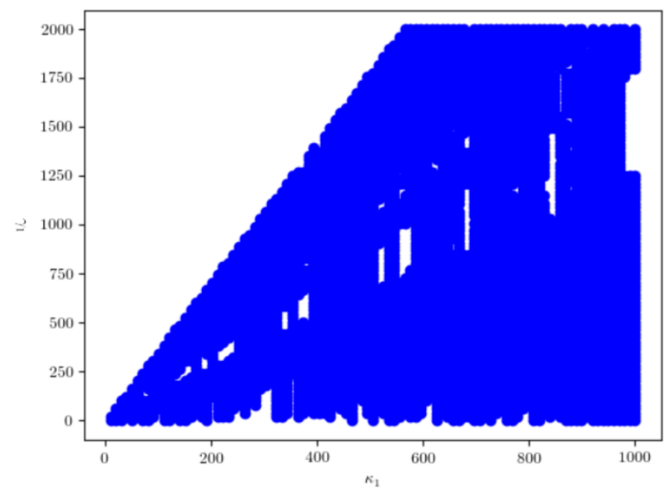
## Linear ERK Model:




boundary of confidence regions for wild-type and all mutants for $\alpha = 0.05$

⇝ practically identifiable

## Rational ERK model:




marginalised confidence area for $\alpha = 0.05$ for the wild-type data point with $0 < \gamma_1 < 1000$ and $0 < \kappa_1 < 1000$, computed numerically. ⇝ practically non-identifiable

Papers:

Eyan Yeung, Sarah McFann, Lewis Marsh, Emilie Dufresne, Sarah Filippi, Heather Harrington, Stanislav Shvartsman, Martin Wühr
Inference of Multisite Phosphorylation Rate Constants and their Modulation via Pathogenic Mutations, Current Biology, Vol 30, 2020

Lewis Marsh, Emilie Dufresne, Helen Byrne, Heather Harrington
Algebra, Geometry and Topology of ERK kinetics, Bulletin of Mathematical Biology 84 (2022)
arXiv: 2112.00688

Thanks for your attention!