# Learning From Data – Model Quality Revisited

**Raymond A. de Callafon**

Prof. at University of California San Diego

Co-founder PXiSE Energy Solutions

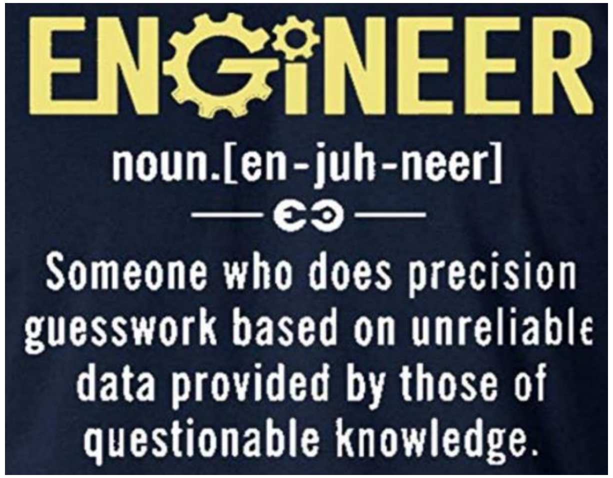
Co-PI NSF funded DERconnect

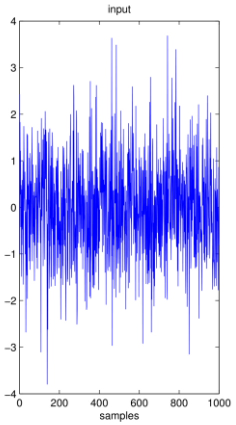
# Why am I here?

- M.Sc student, advised by **Van den Hof**, TUD, 1992 (actuation mechanism in CD player)
- PhD student, **Bosgra** & **Van den Hof**, TUD, 1998 (coprime factor identification, 3D wafer stepper)
- Currently teaching courses on
  - Signals, Systems & Control (undergraduate course & laboratory)
  - Estimation (graduate course on System & State Identification)
  - Control (graduate course on Optimal & Robust control)
- Influenced by Van den HOF (Hall of Famer)

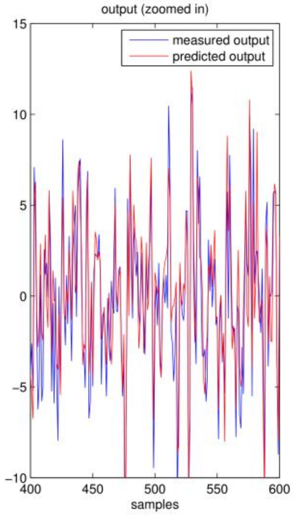
# It is an honor and pleasure to be here!

- **Prof. Van den Hof** has always motivated & inspired (work together to improve science & understanding)
- Developing a dynamic model from data:
  - Intriguing (from *randomness* to *order*)
  - Applicable (dynamic systems *too complex to model*)
  - Gateway (decisions, control, tuning, prediction, …)



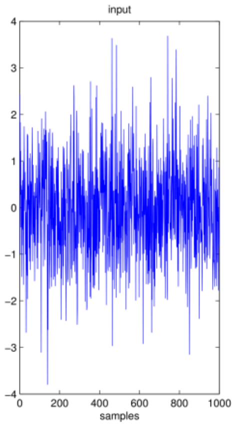

$$u(t) \rightarrow G(q,\theta_0) \rightarrow y(t)$$
$$e(t) \rightarrow H(q,\theta_0) \rightarrow v(t)$$



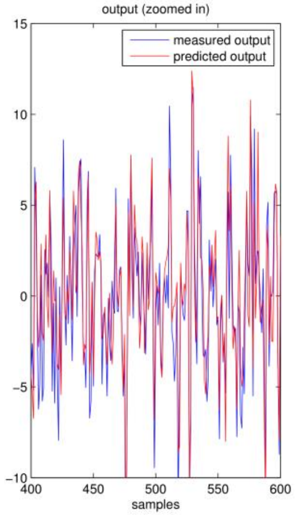

$$\hat{y}(t|t-1,\hat{\theta}) = H(q,\hat{\theta})^{-1}G(q,\hat{\theta})u(t) + (1 - H(q,\hat{\theta})^{-1})y(t)$$

# It is an honor and pleasure to be here!

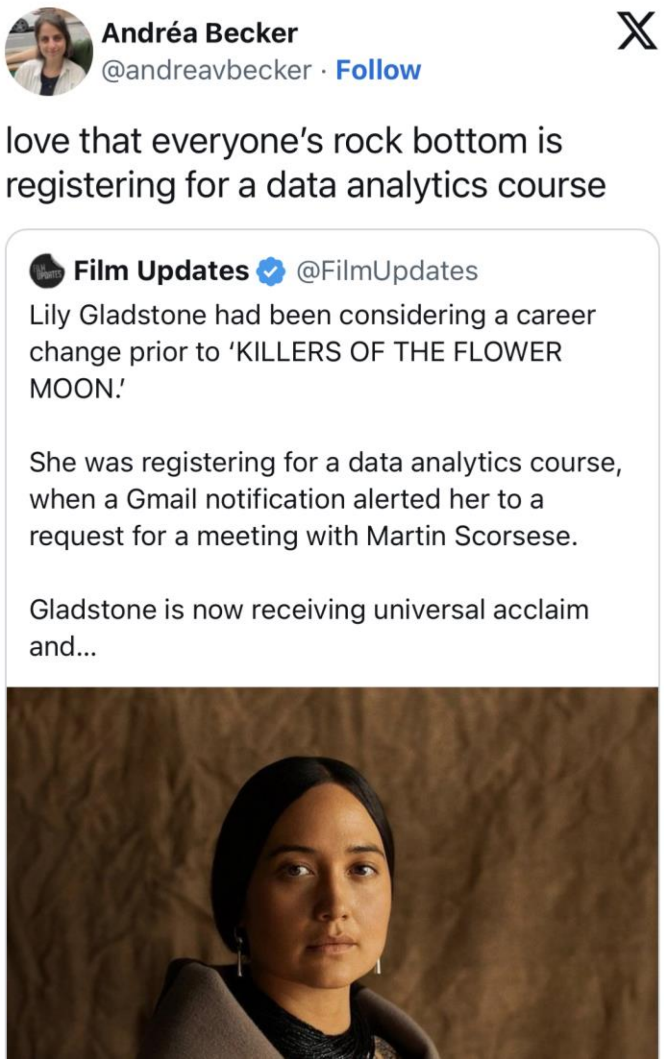
- **Prof. Van den Hof** has always motivated & inspired (work together to improve science & understanding)
- Developing a dynamic model from data:
  - Intriguing (from *randomness* to *order*)
  - Applicable (dynamic systems *too complex to model*)
  - Gateway (decisions, control, tuning, prediction, …)




$$u(t) \rightarrow G(q, \theta_0) \rightarrow y(t)$$
$$e(t) \rightarrow H(q, \theta_0) \rightarrow v(t)$$

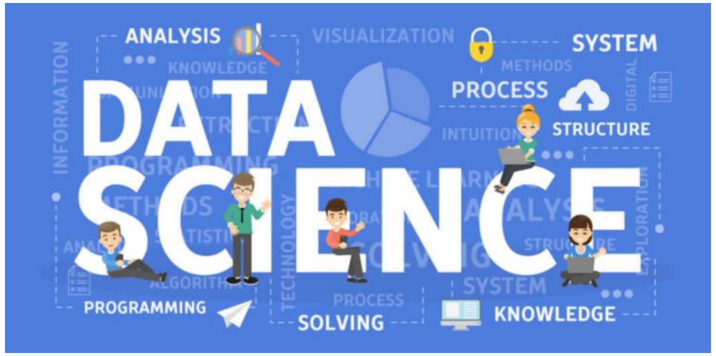



$$\hat{y}(t|t-1, \hat{\theta}) = H(q, \hat{\theta})^{-1} G(q, \hat{\theta}) u(t) + (1 - H(q, \hat{\theta})^{-1}) y(t)$$




UC San Diego
JACOBS SCHOOL OF ENGINEERING

# Main takeaway



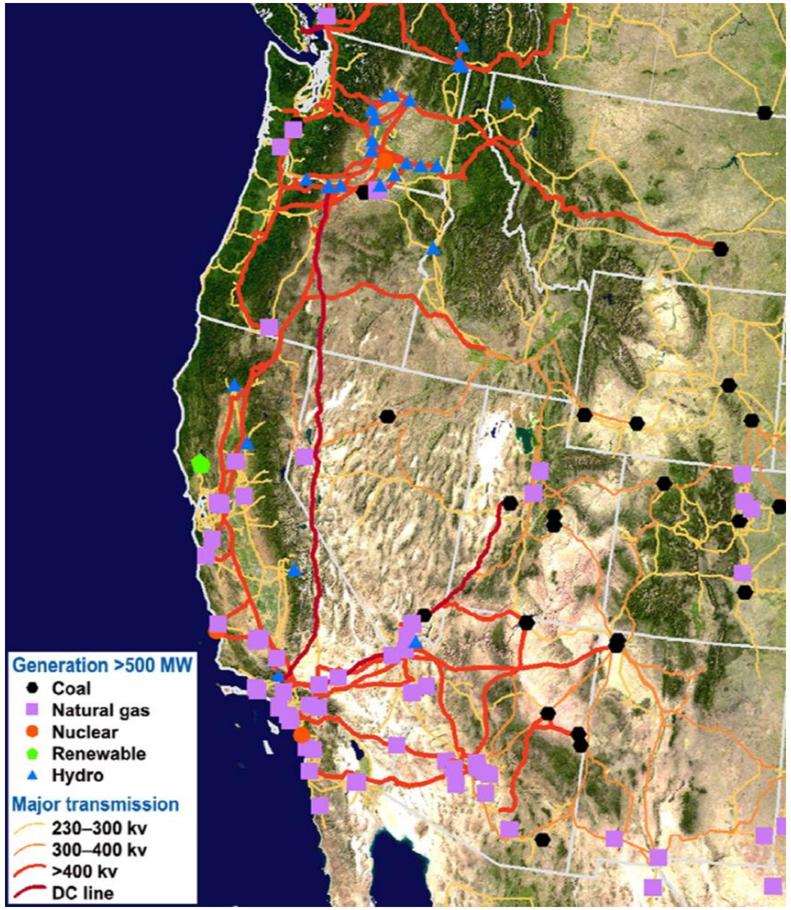

- Exploding activities in Data Science as more data & tools are driving models & decisions
- Data-based modeling has *limitations*:
  - Informative (Persistence of Excitation)
  - Approximate (Lossy Compression: large *N* to little *n*)
  - Noise (Variability & *Bias*)
- Inference (e.g. educated guess) is limited by *model quality*
  - Quantify model quality/error and possibly adjust (application driven)
  - Do not "blindly" depend on model: know its limitations (bias/variability)

*Training a neural network: it knows everything, but you do not know anything (R. Kosut)*

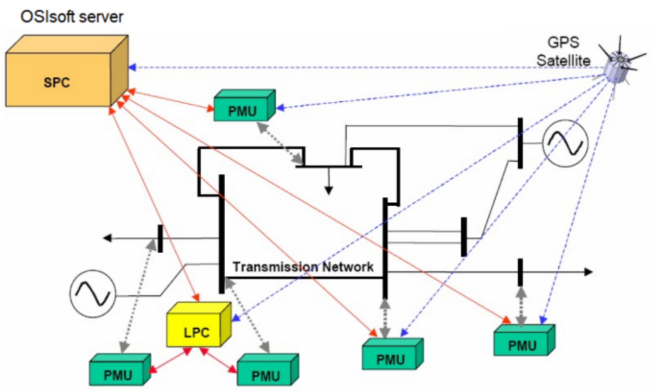
# Quick Outline

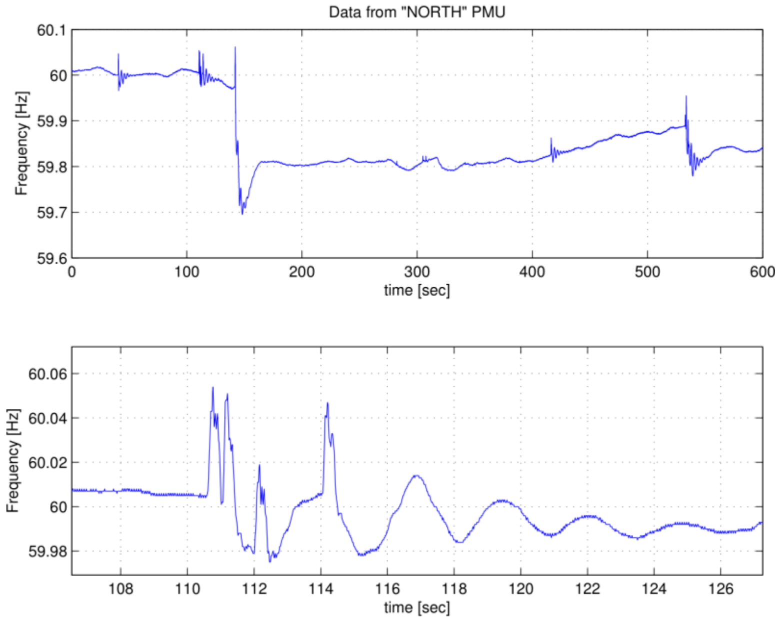
- Illustrations on model quality
  - Practice
  - Theory
- How to capture model quality
  - Inspirations from Robust Control
  - Illustration
- Integration with (robust) control
  - Uncertainty set of models
  - Examples
- Final Recommendations & Remarks

# Illustration of application-driven model quality

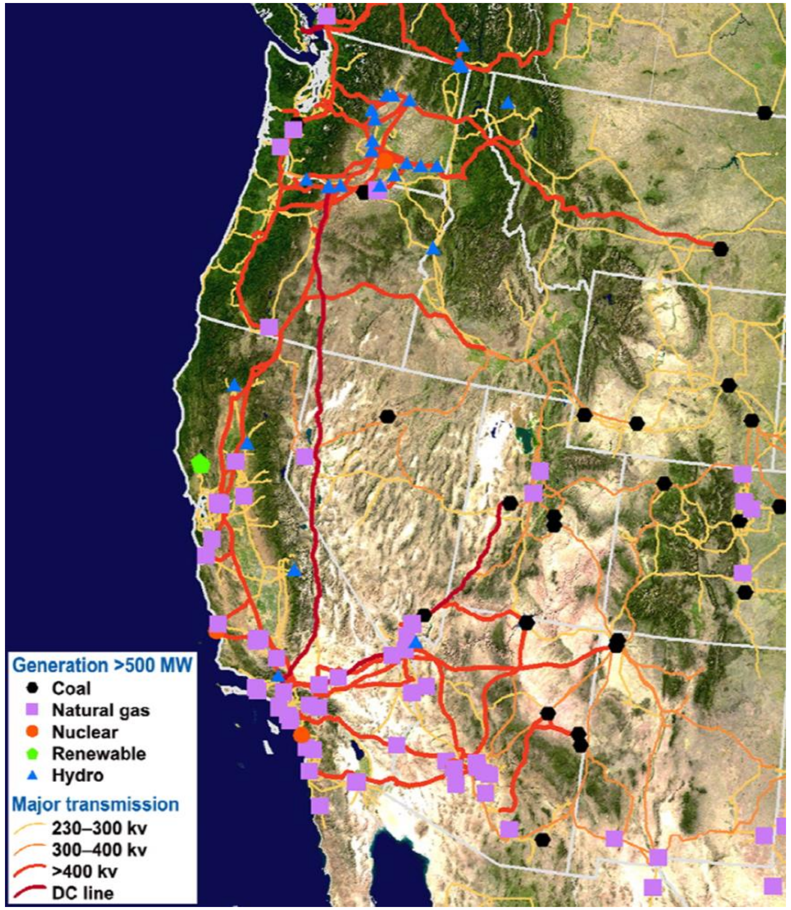



Data: AC frequency and RMS voltage from forced power disturbance (pulse) at Chief Joesph Substation (Brake Test)



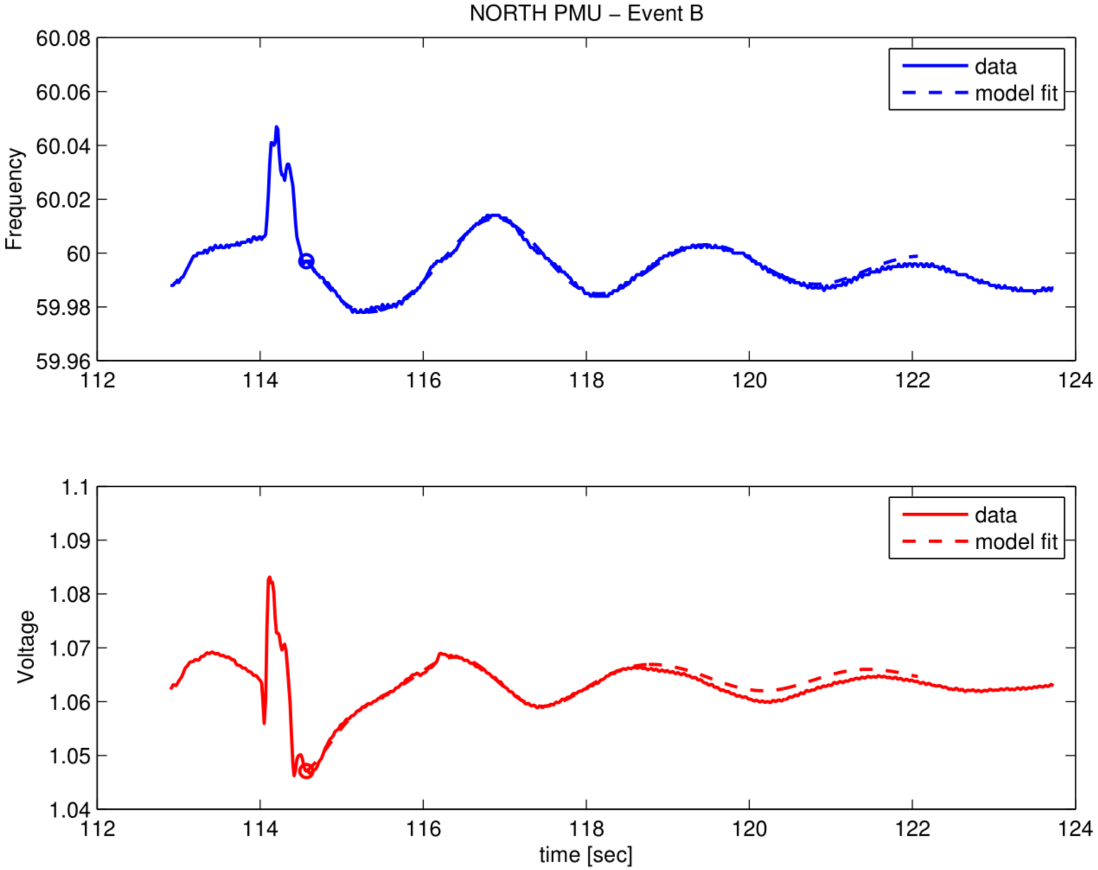
# Illustration of application-driven model quality




$F_1 = 0.38\text{Hz}$
$\beta_1 = 0.0938$

$F_2 = 0.67\text{Hz}$
$\beta_2 = 0.1608$

No additional resonance modes or frequencies!



# Illustration of noise-induced model bias

Noisy data $\{u(t), y(t)\}$ for $t = 1,2, \dots, N$ is given by:

$$y(t) = b_0 \cdot u(t-1) - a_0 \cdot y(t-1) + w(t)$$

Model with $\theta = [b \quad a]^T$ (with $n = 2$) is given by:

$$y(t) = b \cdot u(t-1) - a \cdot y(t-1) + \varepsilon(t, \theta)$$

"Optimal parameter" via LS is related to $R_{yu}(\tau) = E\{y(t)u(t-\tau)\}$:

$$\theta^\star = \lim_{N\to\infty} \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t,\theta)^2 = \begin{pmatrix} R_u(0) & -R_{yu}(0) \\ -R_{yu}(0) & R_{y(0)} \end{pmatrix}^{-1} \begin{pmatrix} R_{yu}(1) \\ -R_y(1) \end{pmatrix}$$

We hope (and expect): $\theta^\star = [b_0 \quad a_0]^T$.

# Illustration of noise-induced model bias

**Unfortunately**, very likely $\theta^\star \neq [b_0 \quad a_0]^T$ as $w(t)$ in

$$y(t) = b_0 \cdot u(t-1) - a_0 \cdot y(t-1) + w(t)$$

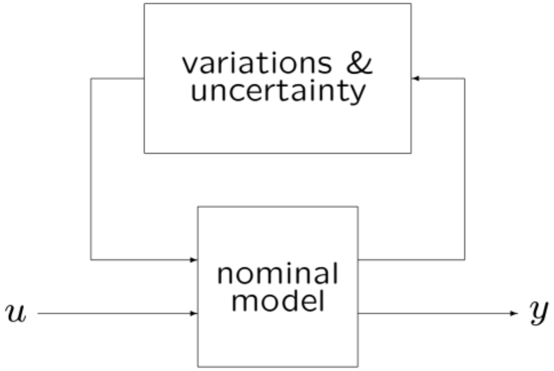
is “not white noise”.

In fact: $w(t) = e(t) + c_0 \cdot e(t-1)$, then

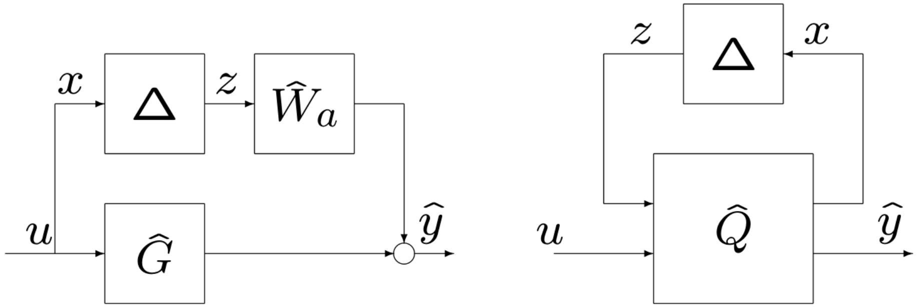
$$\theta^\star = \begin{pmatrix} b_0 \\ a_0 - c_0/R_y(0) \end{pmatrix}$$

Possible solutions: use different model/method, instrumental variables

# How do we capture model quality?

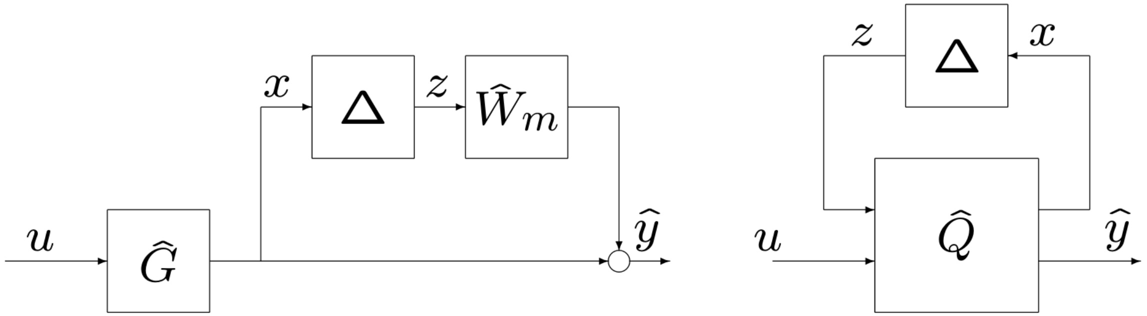



- **We are lucky**: robust control provides a framework
- Additive uncertainty $\Delta_a$, where $\Delta_a = \widehat{W}_a \Delta, ||\Delta||_\infty < 1$:



$$\begin{aligned} x &= 0z + 1u \\ \hat{y} &= \widehat{W}_a z + \widehat{G} u \end{aligned}$$

- Multiplicative uncertainty $\Delta_m$, where $\Delta_m = \widehat{W}_m \Delta, ||\Delta||_\infty < 1$:

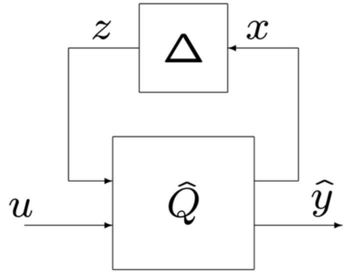


$$\begin{aligned} x &= 0z + \widehat{G} u \\ \hat{y} &= \widehat{W}_m z + \widehat{G} u \end{aligned}$$

# How do we capture model quality?

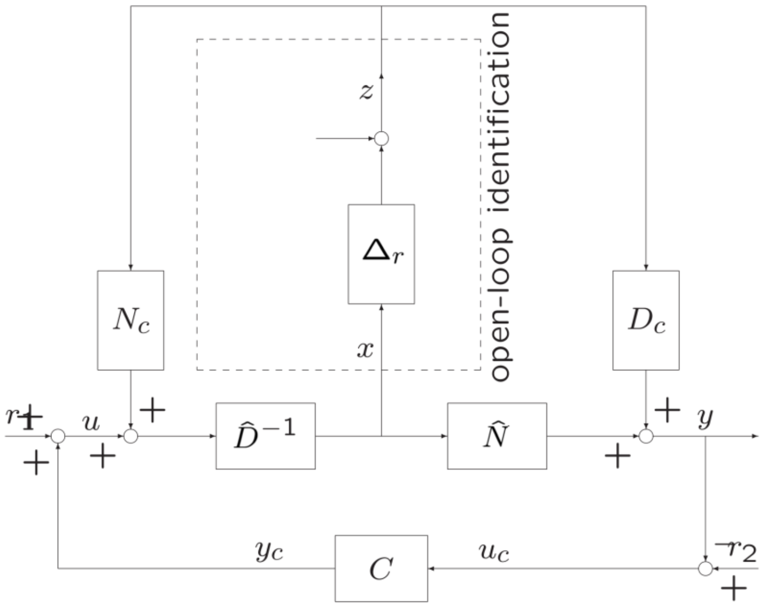
- Especially interesting for models used in control: dual-Youla uncertainty $\Delta_r$ with

$$G = (\hat{N} + D_c\Delta_r)(\hat{D} - N_c\Delta_r)^{-1},\ \Delta_r = \hat{W}_r\Delta,\ \|\Delta\|_\infty < 1$$




- Favorable properties:
  - Uses (exisiting) stabilizing controller $C = N_c D_c^{-1}$
  - All models stabilized by $C$ iff $\Delta_r$ stable
  - Map $\Delta_r$ accessible via signals $x, z$ (filtering)
  - Equivalent open-loop identification

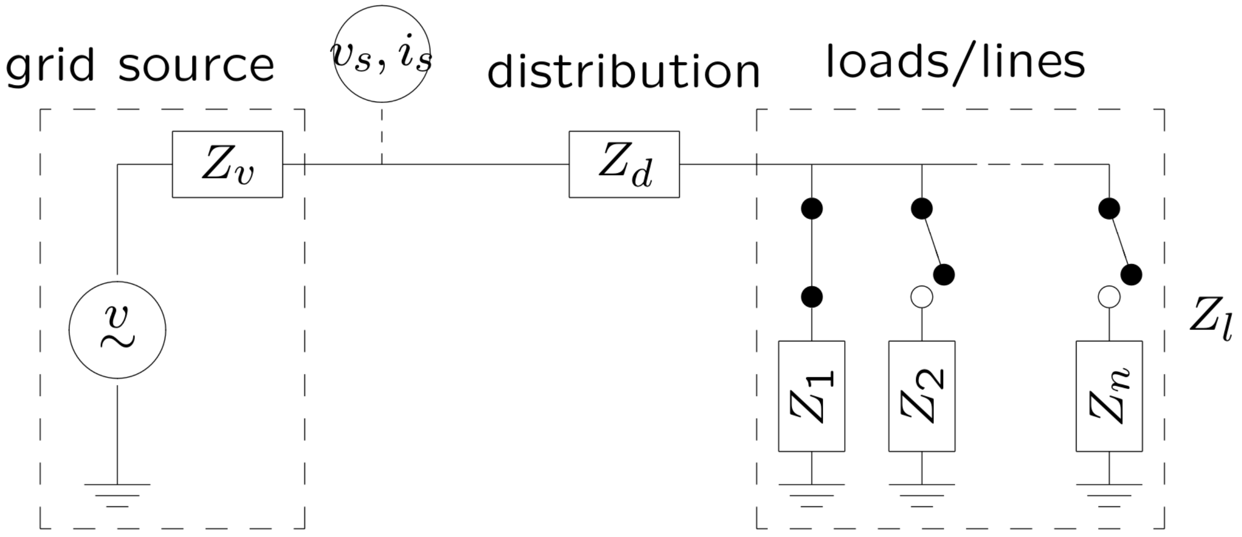
$$\begin{bmatrix} \hat{Q}_{11} & \hat{Q}_{12} \\ \hat{Q}_{21} & \hat{Q}_{22} \end{bmatrix} = \begin{bmatrix} \hat{D}^{-1}N_c\hat{W}_r & \hat{D}^{-1} \\ (D_c + \hat{G}N_c)\hat{W}_r & \hat{G} \end{bmatrix}$$

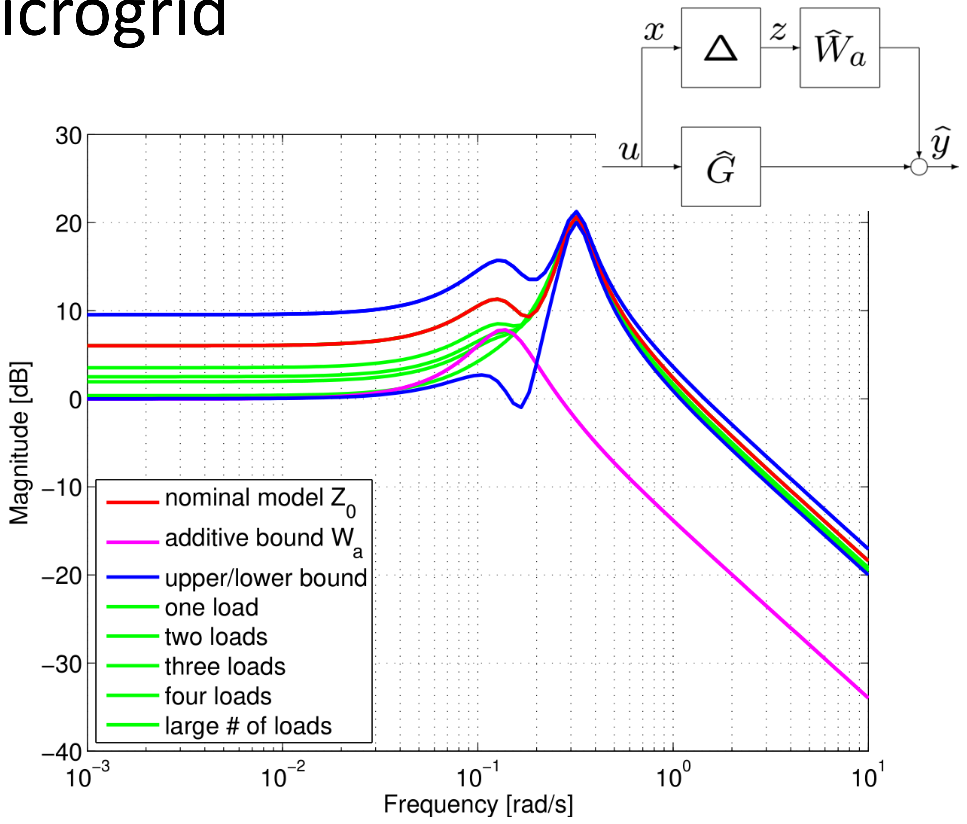


# How do we capture model quality?

**Illustration**: variable inductive loading in microgrid

Load modelled as additive uncertainty $\widehat{W}_a$



# Integration with (robust) control/decisions

- **Inference** using modelled data is limited by quality of model
- Required **quality of model** dictated by application of model
- What if application of model is **control**?
  - Control provides robustness against model errors
  - Model with its errors must allow for control
- Examples for a stable model $\hat{G}$ (and stable $C$) and new $C_{new}$:
  - Additive: $\left\| \hat{W}_a \cdot \frac{C_{new}}{1+C_{new}\hat{G}} \right\|_\infty < 1$, hence: $\hat{W}_a$ small, where $\frac{C_{new}}{1+C_{new}\hat{G}}$ large
  - Dual-Youla: $\left\| \hat{W}_r \cdot \frac{(C-C_{new})}{1+C_{new}\hat{G}} \right\|_\infty < 1$, hence: $\hat{W}_r$ small where $\frac{(C-C_{new})}{1+C_{new}\hat{G}}$ large

# Integration with (robust) control/decisions

- With model quality characterized by an uncertainty set of models:

$$\mathcal{G} = \{G \mid G = \mathcal{F}(\widehat{G}, \Delta),\ \ \|\Delta\|_\infty < 1\}$$

$$\mathcal{F}(\widehat{G}, \Delta) = [\widehat{N} + D_c\Delta_r][\widehat{D} - N_c\Delta_r]^{-1},\ \ \Delta_r = W_r\Delta$$

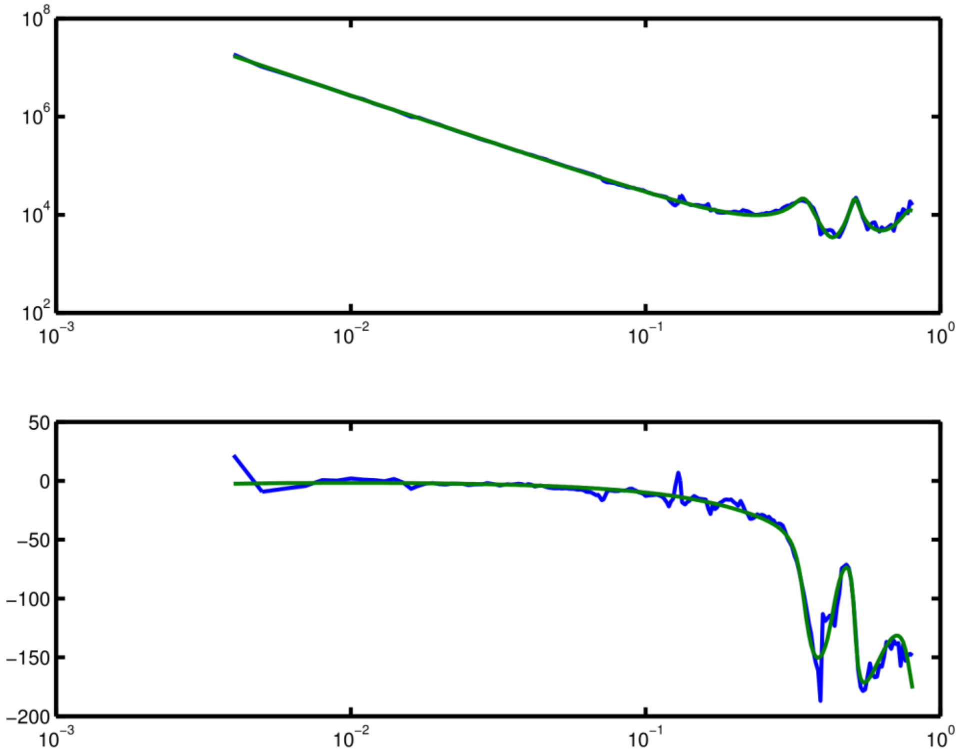
- With update/new control $C_{new}$, quality $\widehat{W}_r$ **must satisfy**

$$\left\| \widehat{W}_r \cdot \frac{(C - C_{new})}{1 + C_{new}\widehat{G}} \right\|_\infty < 1$$

- Shows interplay between **model quality** and **control application**
- Model: **control-relevant** or **“closed-loop”** (also: Δ inequality)

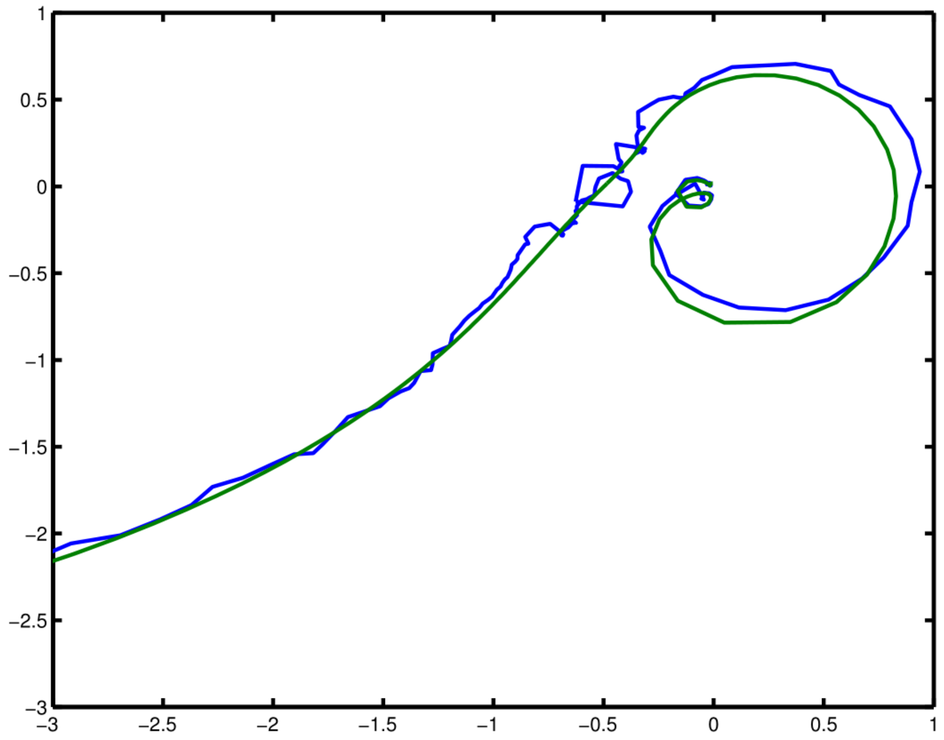
# Integration with (robust) control/decisions

- **Illustration**: frequency domain data of CD actuation mechanism




frequency data ■

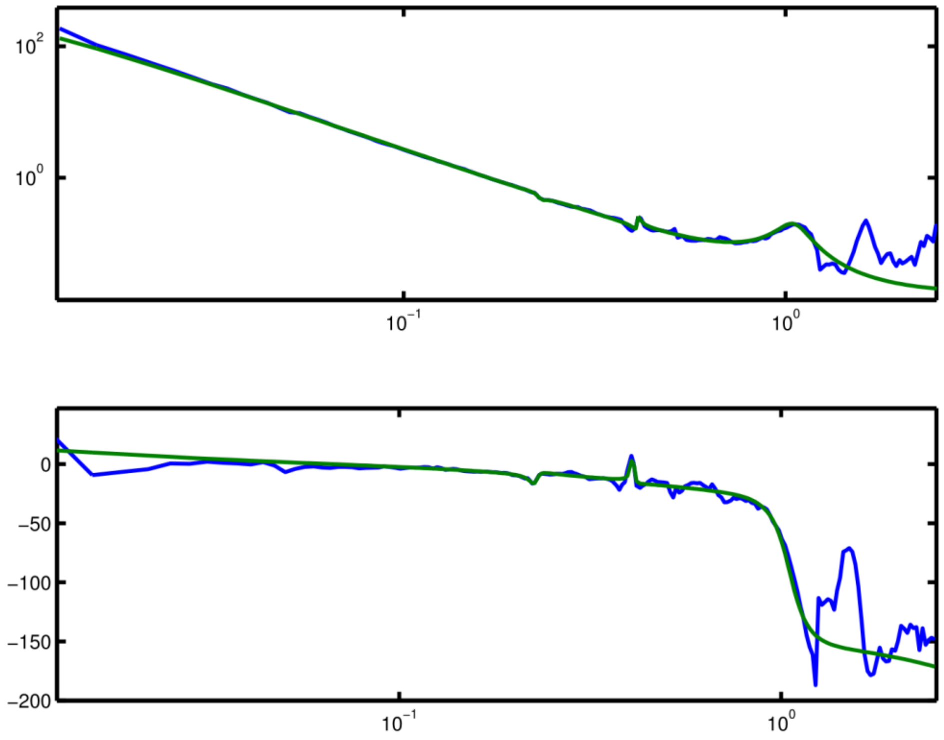
frequency response ■ "open-loop" model




UC San Diego
JACOBS SCHOOL OF ENGINEERING

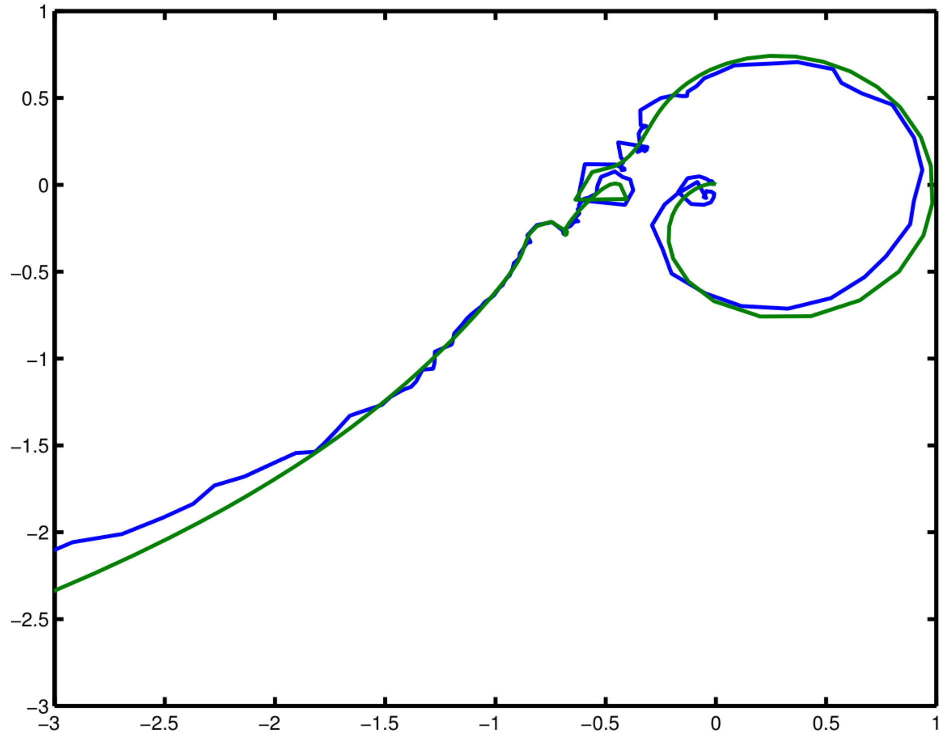
# Integration with (robust) control/decisions

- **Illustration**: frequency domain data of CD actuation mechanism




frequency data ■

frequency response ■ "closed-loop" model




UC San Diego
JACOBS SCHOOL OF ENGINEERING

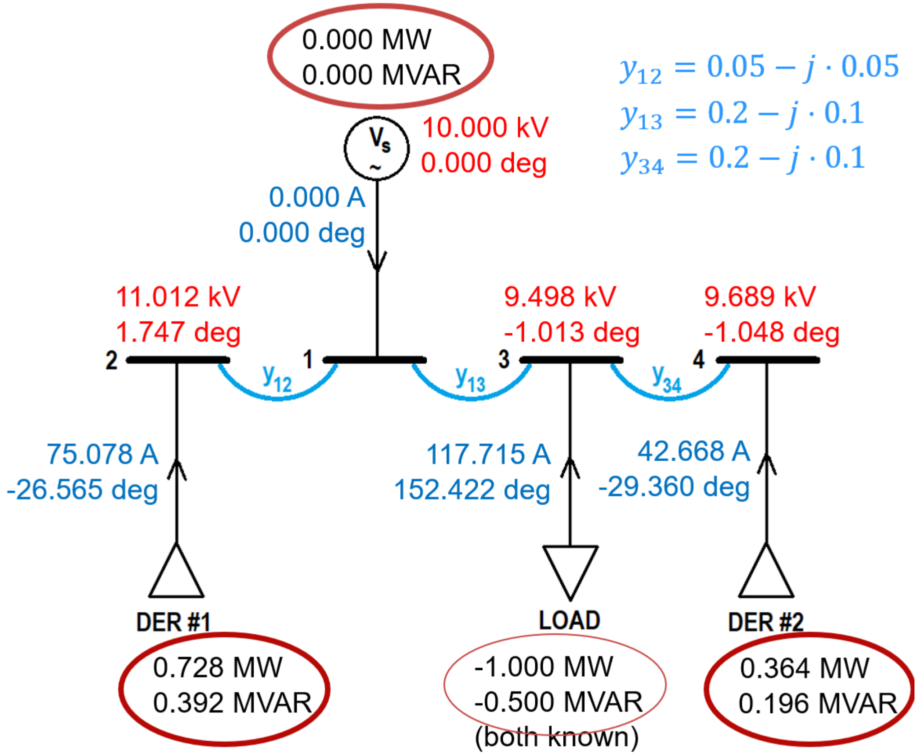
# Integration with (robust) control/decisions

**Illustration:** Power System

Let DER #1 = 2 x DER #2. What P,Q should we assign to DERs so that Power Source is 0 MW, 0 MVAR? (think: POI or generator)

DER power assignment (voltage and current) can be solved with *full network model/data* (load data + admittance data $y_{12}, y_{13}, y_{34}$)

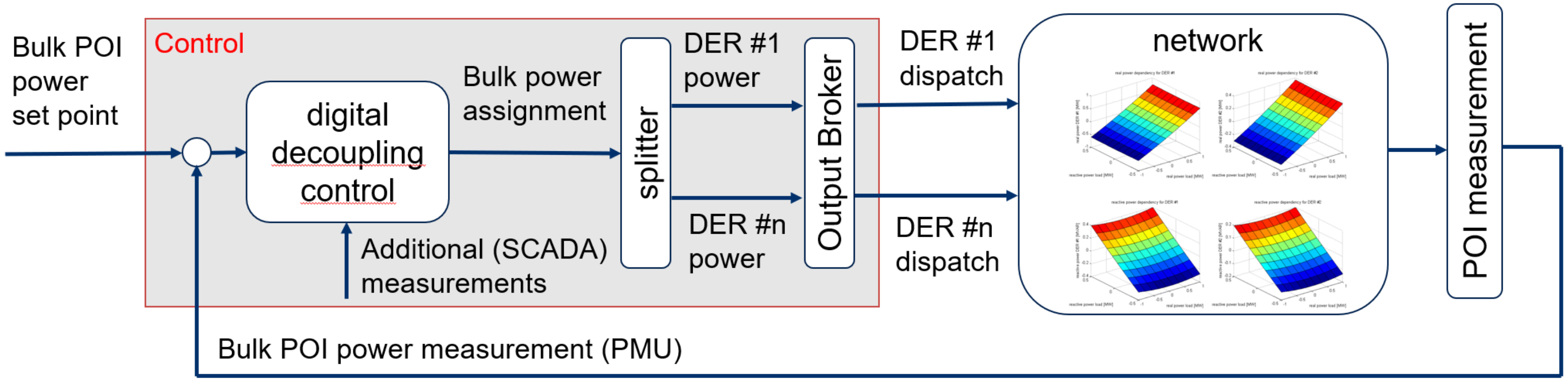
- Realistic to assume access to full load and admittance information?
- Alternative solution with simple model and control?

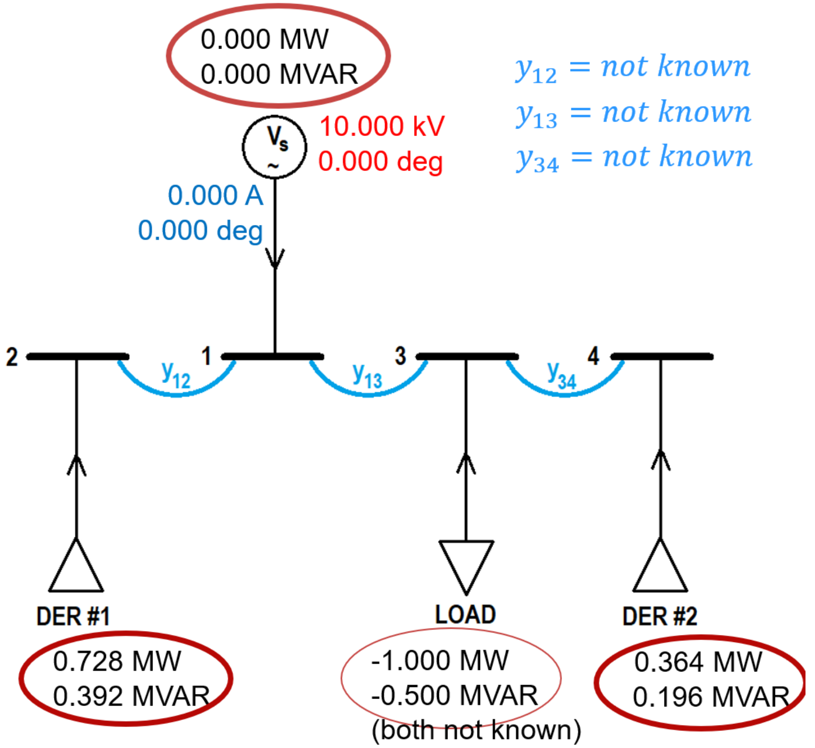


# Integration with (robust) control/decisions

**Solution** with simplified model + control:

- Only connectivity, no impedance/load information (very low-quality model)
- Increase of bulk P of DERs: increase of P at POI
- Increase of bulk Q of DERs: increase of Q at POI
- DER dynamics: time delay + settling



# Recommendations (formal)

Data-based modeling for control should consists of two steps:

1. IDENTIFICATION: Estimate new uncertainty set $\mathcal{G}^{i+1}$ ($\mathcal{G}^{i+1} = \{G \mid G = \mathcal{F}(\widehat{G}^{i+1}, \widehat{W}_{\Delta}^{i+1})\}$ with $G_0 \in \mathcal{G}^{i+1}$) via

$$(\widehat{G}^{i+1}, \widehat{W}_{\Delta}^{i+1}) = \arg \min_{G, W_{\Delta}} \sup_{G \in \mathcal{G}^{i+1}} J(G, W^i, C^i)$$

such that the worst-case performance has improved:

$$\sup_{G \in \mathcal{G}^{i+1}} J(G, W^i, C^i) \leq \sup_{G \in \mathcal{G}^i} J(G, W^i, C^i) < 1$$

and is a model validation test before using the using the new uncertainty set $\mathcal{G}^{i+1}$ for control design.

UC San Diego
JACOBS SCHOOL OF ENGINEERING

# Recommendations (formal)

Data-based modeling for control should consists of two steps:

2. CONTROL DESIGN: compute new robust controller

$$C^{i+1} = \arg\min_{C} \sup_{G \in \mathcal{G}^{i+1}} J(G, W, C)$$

such that (again) the worst-case performance has improved:

$$\sup_{G \in \mathcal{G}^{i+1}} J(G, W^i, C^{i+1}) \leq \sup_{G \in \mathcal{G}^{i+1}} J(G, W^i, C^i)$$

and is a controller validation test before implementing the new controller $C^{i+1}$ on the plant $G_0$.

# Concluding Remarks

Lessons from System Identification with Robust Control:

- Parametrize model quality/uncertainty
- Estimate models not by just “fitting data”
- Let data reduce model uncertainty relevant for the application
- Modeling from data is an iterative process (reinforcement)
- Applicable to data science & data-based modeling

Thank You