# Autotuner identification of TITO systems using a single relay feedback experiment

Berner, Josefin; Soltesz, Kristian; Hägglund, Tore; Åström, Karl Johan

# Autotuner identification of TITO systems using a single relay feedback experiment

**Josefin Berner, Kristian Soltesz, Tore Hägglund, Karl Johan Åström** [*]

[*] *All authors are with Department of Automatic Control, Lund University, Sweden. Contact* `josefin.berner@control.lth.se`

**Abstract:** Relay autotuning has proven very successful for single-input single-output systems. This paper proposes an identification method for relay autotuning of systems with two inputs and two outputs (TITO systems). The combination of asymmetric relay feedback and output error identification admits short tuning time, without the need for limit cycle convergence. The method is successfully demonstrated on relevant system models, including the Wood-Berry distillation column.

*Keywords:* Multivariable autotuning, decentralized relay experiment, output error identification

## 1. INTRODUCTION

Relay autotuning for single-input single-output (SISO) systems has been widely used since its introduction in the 1980's. The closed-loop identification automatically excites the system in the frequency interval relevant for PID control, and the amplitudes of the oscillations can be kept small as to not disturb the process more than necessary. The relay autotuner in Åström and Hägglund (1984) has been modified by e.g. Luyben (1987) where a first-order model with time delay (FOTD model) of the system was derived, by Shen et al. (1996) where an asymmetric relay was used, and in more recent work as Berner et al. (2016b) where the normalized time delay was used to classify the system during tuning and Berner et al. (2016a) where parameter choices and other practical issues were discussed. A recent review of the advances in identification from relay experiments is given in Liu et al. (2013). Several books on relay autotuning, like Yu (2006); Chidambaram and Sathe (2014), also give good overviews.

Even though much has been written about relay autotuning almost all of it considers SISO systems only. Since many industrial processes are of multi-input multi-output (MIMO) type, there is a need of finding multivariable models of them. These models can then be used to either tune a number of SISO PID controllers by picking a suitable input-output pairing or by decoupling the system. Or a multivariable PID controller could be tuned as in Boyd et al. (2015). The choice of tuning method is not the subject of this paper, but by obtaining a full transfer function matrix, many tuning options are possible. In this paper we will restrict ourselves to two-input two-output (TITO) systems. This is a common subclass of MIMO

systems that shows up in many places both in literature and in industrial processes.

Relay autotuning of multivariable systems could be done in three different ways, Wang et al. (1997). The first way is to tune each loop independently while leaving the others in manual. This method does not take any of the cross-couplings in the system into account, and hence is not a good option for coupled systems. The second method is sequential tuning, where the first loop is tuned while the others are in manual, and then the next loop is tuned with the first one closed, and so on until all loops are tuned. With this method the controllers are tuned based on all the information up to that point, but the loops that are still open do not influence the behavior. Therefore the method is usually iterated a number of times with the loops closed, which results in a total of $mk$ relay experiments for $k$ iterations of an $m \times m$ process. The third option, which we will use in this paper, is decentralized relay feedback, where all the loops are tuned simultaneously. This is a completely closed-loop method where all cross-couplings will influence the result. In Wang et al. (1997) $m$ decentralized relay tests were needed in order to obtain the transfer functions. It was also needed to wait for convergence of the limit cycle oscillations. Another drawback of that method is the assumption that the systems are coupled strongly enough to oscillate with the same fundamental period. Attempts to find the conditions for when this happens has been made in e.g. Loh and Vasnani (1994), but to our knowledge no simple full conditions have been published. In Campestrini et al. (2006), problems with using the ultimate frequency as a tuning parameter for MIMO systems have been remarked upon. For MIMO systems there is not an ultimate point as in the SISO case, but rather an ultimate surface, and which point that will be found from the experiment depends on the settings. This makes simple methods like
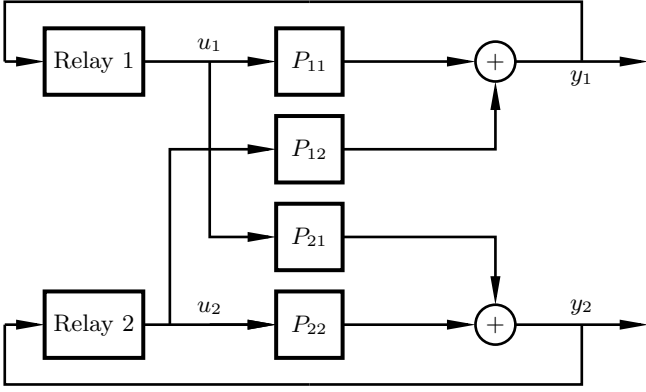
Fig. 1. Setup of the decentralized relay experiment.

Ziegler-Nichols, based on the ultimate gain and frequency, inappropriate for the MIMO case.

What we want is a relay autotuner for MIMO systems that is fast, does not require numerous experiments, that works for both weakly and strongly coupled processes, and that does not rely on only an ultimate point to get the process models. The autotuner in this paper aims to identify the transfer function matrix for a TITO system from one single decentralized relay experiment. It does not need to wait for convergence, which makes it fast. The models are estimated by output error minimization, which does not require the loops to oscillate with the same frequency and does not use the ultimate frequency in any calculations. The method will be described and some examples demonstrating the potential of the method will be given. The controller tuning, and some practical issues of the experiments are left as future work.

## 2. METHOD

### 2.1 Experiment

The experimental setup is shown in Fig. 1 and based on the decentralized relay feedback experiment proposed in Wang et al. (1997). Both loops are closed with relay feedback simultaneously. Even though it is a closed-loop experiment, the input-output noise correlation is negligable since the input signal is kept constant except at the switching instances of the relay. Hence, the system can be viewed as open-loop for identification purposes. The relays used are asymmetric, and implement most practical features from the one described in Berner et al. (2016a). To get as much excitation as possible the asymmetry level $\gamma$ (i.e. the ratio between the on and off amplitudes) is different for the two relays. In this paper the asymmetries were set to $\gamma = 1.5$ in relay 1 and $\gamma = 2$ in relay 2, but this choice should be investigated more in future work. The amplitudes are set automatically during the startup phase. The identification method that will be used in this paper does not require limit cycle convergence, hence the experiments can be made short. In this paper the experiments are stopped when both loops have undergone four relay switches. For the simulations in this paper, the control signal is set to its stationary level immediately when the experiment stops. If this is the best way of shutting down the experiment, and when to connect the new controller, needs to be investigated in future work.
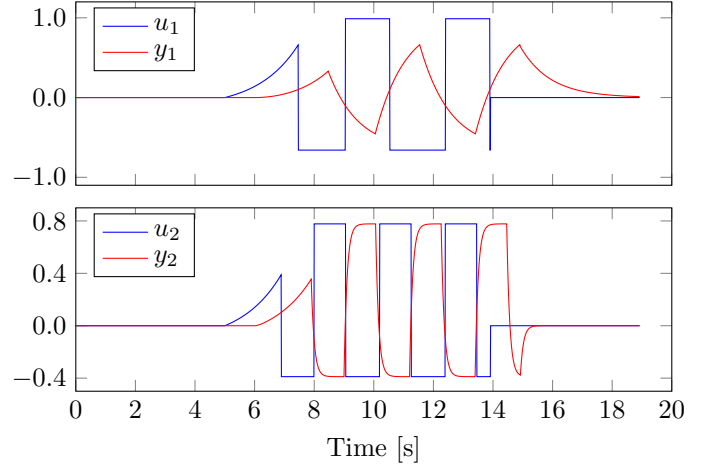


Fig. 2. Decentralized relay experiment for the uncoupled system, $G_{\epsilon=0}$. The two loops converge to limit cycles with different periods $t_p = 3.26$ and $t_p = 2.19$.

### 2.2 Coupling level

As has been described in Loh and Vasnani (1994), there are three possible limit cycle behaviors of the decentralized relay feedback experiment of the TITO system. The separate behaviors depend on the strength of the cross-couplings in the process. An illustration of the possible outcomes is given by looking at the system

$$G_\epsilon(s) = \begin{pmatrix} \dfrac{1}{s+1}e^{-s} & \dfrac{\epsilon}{s+1}e^{-s} \\ \dfrac{\epsilon}{s+1}e^{-s} & \dfrac{1}{0.1s+1}e^{-s} \end{pmatrix} \quad (1)$$

with $\epsilon = \{0, 0.3, 0.8\}$ shown in Fig. 2-4. For the purpose of this illustration the decentralized relay experiment has been run until convergence of the time periods of the limit cycles, or in the intermediately coupled case for 10 switches since it will not converge. If the coupling of the system is sufficiently strong, like in Fig. 4, both loops will oscillate with the same frequency, the half-periods may, however, be separate. If the coupling is weak (or non-existent), as in Fig. 2, the loops will oscillate with different frequencies like two separate SISO systems. If the coupling is intermediate, see Fig. 3, the loops will have a more complex limit cycle with multiple relay switches within the fundamental period.

Most literature on decentralized relay autotuning like e.g. Wang et al. (1997) assumes the strongly coupled case with equal fundamental periods. We would, however, like a method that can handle all coupling levels. The uncoupled system could just as well be treated as two SISO systems and use methods for that, but the possibility to handle it in the same framework as the others is desirable since the coupling level is not necessarily known on beforehand.

Fig. 3 and Fig. 4 show that the waveforms of the limit cycles can be quite complex, and to use simple equations from a few significant data points, as is usually done in the SISO autotuner, is not straightforward. Instead we use all data points to identify model parameters, as described in the following section.
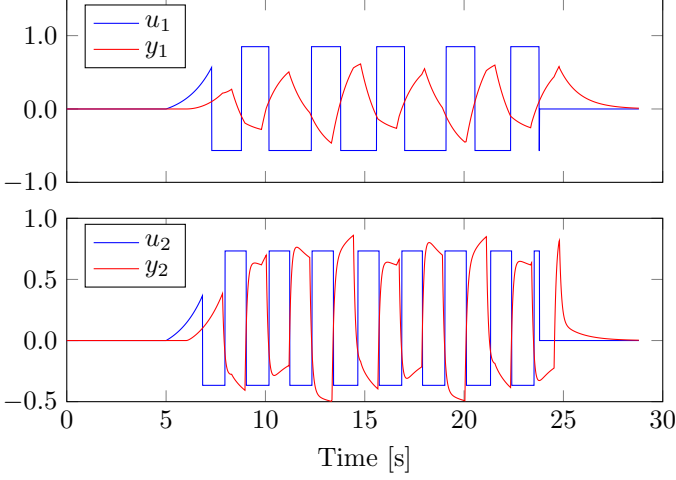
Fig. 3. Decentralized relay experiment for the intermediately coupled system, $G_{\epsilon=0.3}$. The loops converge to limit cycles that contain multiple relay switches.
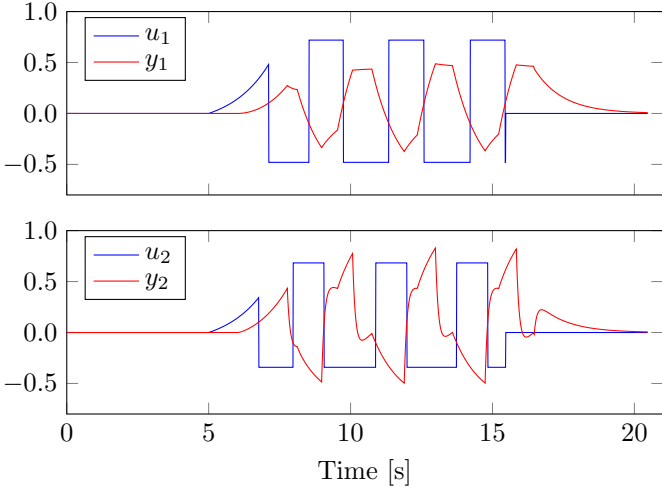


Fig. 4. Decentralized relay experiment for the strongly coupled system, $G_{\epsilon=0.8}$. The loops converge to limit cycles with identical periods $t_p = 2.76$.

*2.3 Identification method*

The elements of the transfer function matrix in Fig. 1 will be estimated on the form

$$P_{i,j}(s) = \frac{b_{i,j}}{s + a_{i,j}} e^{-sL_{i,j}}, \ i,j \in \{1,2\}, \quad (2)$$

that is, first order systems with time delay (FOTD). The parameter vector defining each element (2) is $\theta_{i,j} = [b_{i,j} \ a_{i,j} \ L_{i,j}]^\top$.

For our purposes the system can be decomposed into two parts $y_i = P_{i,1}u_1 + P_{i,2}u_2$, with parameter vectors $\theta_i = \begin{bmatrix} \theta_{i,1}^\top & \theta_{i,2}^\top \end{bmatrix}^\top$. Denoting by $\mathbf{u}_1, \mathbf{u}_2, \mathbf{y}_i$ and $\hat{\mathbf{y}}_i$, the sampled process inputs, output, and the model output, respectively, we pose the parameter identification problem as minimization of the $\mathcal{L}_2$ norm of the output error $\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i$

$$J(\theta_i) = \frac{t_s}{2}\mathbf{e}_i^\top \mathbf{e}_i, \quad (3)$$

where $t_s$ is the sampling period of the signals. Since $\hat{\mathbf{y}}_i$ is not convex in $\theta_i$, a local second-order method will be used. Its SISO counterpart for discrete time systems was described in Åström and Bohlin (1965), and it was adopted to relay identification of continuous time systems in Soltesz et al. (2010). In the SISO case, the essence of the method lies in constructing a continuous time state space system $\mathcal{S} : \{A, B, C, D\}$. The system $\mathcal{S}$ is parametrized in $\theta$, and is designed to have a certain structure such that its delayed output contains both the estimated output $\hat{\mathbf{y}}$ and its gradient $\nabla_\theta \hat{\mathbf{y}}$,

$$\mathrm{III}_{t_s} \ C\mathbf{x}(t + L) = \begin{bmatrix} \hat{\mathbf{y}}^\top & \nabla_\theta \hat{\mathbf{y}}^\top \end{bmatrix}^\top,$$

where $\mathrm{III}_{t_s}$ denotes the sample operator of period $t_s$. The (un-delayed) output of $\mathcal{S}$ is cheap to compute by simulation of the discrete time system $\{\Phi, \Gamma, C, D\}$ obtained through zero-order hold sampling of $\mathcal{S}$, with period $t_s$. A subsequent shift by $\lfloor L/t_s \rfloor$ produces the desired output and associated gradient. For the FOTD SISO case, a minimal realization of $\mathcal{S}$ is given by

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{cc|c} -a & 0 & 1 \\ b & -a & 0 \\ \hline b & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \\ ab & 0 & -b \end{array}\right]. \quad (4)$$

See Soltesz et al. (2010) for the general order SISO case, and its derivation.

Use of (4) enables computation of the Jacobian

$$\nabla_\theta J = t_s \mathbf{e}^\top \nabla_\theta \hat{\mathbf{y}},$$

and an approximation of the associated Hessian

$$\Delta_\theta J \approx \nabla_\theta \hat{\mathbf{y}}^\top \nabla_\theta \hat{\mathbf{y}}.$$

The approximation consists in keeping quadratic terms $\partial^2/\partial^2\theta_i$, but discarding cross terms $\partial^2/(\partial\theta_i\partial\theta_j)$ – which are negligible unless $\mathbf{e}$ and $\nabla_\theta \hat{\mathbf{y}}$ are strongly correlated.

Using the Jacobian and Hessian information, the optimization problem can be solved (to a local minimum) by standard means. (The trust-region-reflective solver bundled with Matlab has been used here.) Each iteration only involves a simulation of $\mathcal{S}$ and a gradient descent step, making the identification method computationally cheap.

Extension to the TITO case is straightforward as the output $y_i = P_{i,1}u_1 + P_{i,2}u_2$ is the sum of outputs $y_{i,1} = P_{i,1}u_i$ and $y_{i,2} = P_{i,2}u_2$. Two instances of (4) are constructed for $y_{i,1}$ and $y_{i,2}$, respectively. From their outputs it is straightforward to assemble $\nabla_{\theta_i}\hat{\mathbf{y}}_i = \begin{bmatrix} \nabla_{\theta_{i,1}}\hat{\mathbf{y}}_i & \nabla_{\theta_{i,2}}\hat{\mathbf{y}}_i \end{bmatrix}$, which enables computation of the Jacobian and Hessian associated with (3).

It can be noted that the identification method is readily extendible to higher order systems, as put forth in Soltesz et al. (2010). It is also straightforward to apply it to transfer matrices of arbitrary input and output dimensionality. The ability of the proposed method to identify parameters under such conditions, depends on how well the identification input excites the dynamics, and will be investigated in future work.

*Excitation* As was just stated, it is crucial to excite the process sufficiently to be able to find all unknown process parameters. In Fig. 5 the power spectra of the input
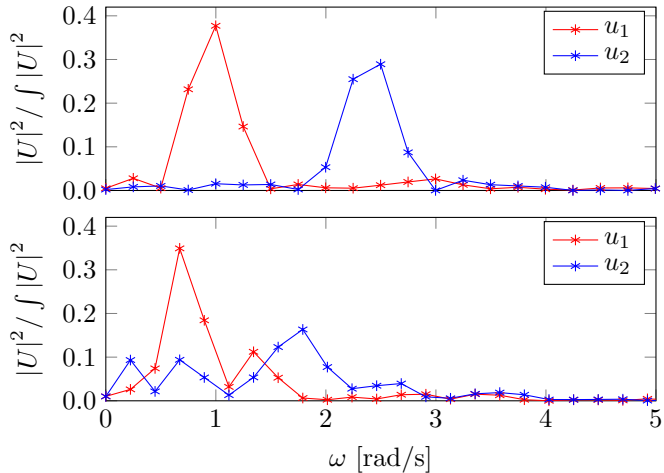
Fig. 5. Power spectra of the relay signals in the decentralized relay experiment for the process $G_{\epsilon=0.3}$. The upper plot shows an experiment with symmetric relays. The lower plot shows an experiment with the asymmetry levels $\gamma_1 = 1.5$ and $\gamma_2 = 2$.

signals are shown for two decentralized relay feedback experiments. These specific experiments were done on $G_{\epsilon=0.3}$, as defined in (1), with symmetric relay functions in the upper plot and asymmetric relay functions in the lower plot. As can be seen the signal power is much more spread out for the asymmetric relay functions. The frequency plots of course depend on the processes. The large peaks in the frequency plots are from the main oscillation periods, so for strongly coupled systems, e.g. $G_{\epsilon=0.8}$, that oscillates with the same frequency, the two curves will more or less be on top of each other.

*Initialization*  The described identification method sometimes converges to a local minimum. Initialization is therefore important. Since the computations are cheap and fast, and this paper is a proof of concept rather than a complete algorithm, we chose to initialize in multiple points and then pick the best solution. The first attempts indicated that the time delay $L$ is the most crucial parameter to have a good initial value of, therefore we start with $L$ gridded between 0 and $L_{max}$ that is here set to the time period of the oscillation. The initial values of $a$ and $b$ are set to 0. If the grid is very dense this methodology will require a lot of combinations which will make the overall computing time larger than necessary. Therefore a more clever way of initializing the system should be found in the future.

## 3. RESULTS

To evaluate the method we explored the three $G_\epsilon$ in (1), representing different coupling levels, and the Wood-Berry distillation column, Wood and Berry (1973). The dynamics of this common benchmark process is given by

$$G_{WB}(s) = \begin{pmatrix} \dfrac{12.8e^{-s}}{1+16.7s} & \dfrac{-18.9e^{-3s}}{1+21s} \\ \dfrac{6.6e^{-7s}}{1+10.9s} & \dfrac{-19.4e^{-3s}}{1+14.4s} \end{pmatrix}. \quad (5)$$

White noise with a peak-to-peak-amplitude of 0.9 was added to the simulations. The experiments for the three $G_\epsilon$
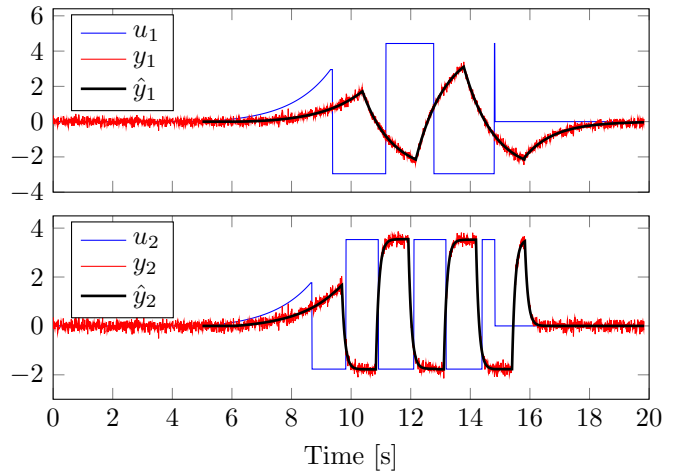


Fig. 6. Experiment data and model fit for the weakly coupled example $G_{\epsilon=0}$.

are shown in Fig. 6-8, and the experiment for Wood-Berry is shown in Fig. 9. The figures show that the black estimated output curves follow the true (noisy) red curves very well. The obtained model parameters are listed in Tab. 1. They are all close to their true values. The estimate of $G_{WB}$ is comparable to the one obtained in Chidambaram and Sathe (2014). Our parameters are slightly worse, but obtained from one noisy experiment instead of two noise-free. Since the method in Chidambaram and Sathe (2014) assumes strong coupling all $G_\epsilon$ cannot be compared.

To this point all examples have been processes that are of the model order we are estimating, and the good accordance of the estimated models is therefore possible. To explore what happens if the process is of higher order we investigated the second order TITO process

$$G_{SOTD}(s) = \begin{pmatrix} \dfrac{1}{(s+1)^2}e^{-s} & \dfrac{0.3}{(s+1)^2}e^{-s} \\ \dfrac{0.3}{(s+1)^2}e^{-s} & \dfrac{1}{(0.1s+1)^2}e^{-s} \end{pmatrix}. \quad (6)$$

The results for this process are shown in Fig. 10 and the Bode plots for the true and estimated models are shown in Fig. 11. There is a very good match between the estimated and the true Bode plots up to frequencies where the phase lag is $-180°$, which means that good PID controllers can be designed based on the FOTD models. The Bode plots naturally differ at high frequencies due to the differences in high frequency roll off. The estimated model parameters for $G_{SOTD}$ are listed in the bottom of Tab. 1. Notice that the estimated time delays $\hat{L}$ are significantly larger than the true time delays $L$. This is because the extra time constant of the SOTD model is split between the time constant and the time delay of the estimated FOTD model. The large difference in true and estimated time delays indicates that, even though a good PID controller for the FOTD model could be obtained, the performance could be improved significantly by basing control design on a model of higher order. This since the time delay causes fundamental limitations on the achievable bandwidth of the system, and a robust control design has a constant product of the time delay and gain crossover frequency.

Table 1. Estimated model parameters $\hat{K}, \hat{T}, \hat{L}$, as well as the true parameters $K, T, L$, for the five example processes. The rows are ordered $P_{11}, P_{12}, P_{21},$ and $P_{22}$ for each process.

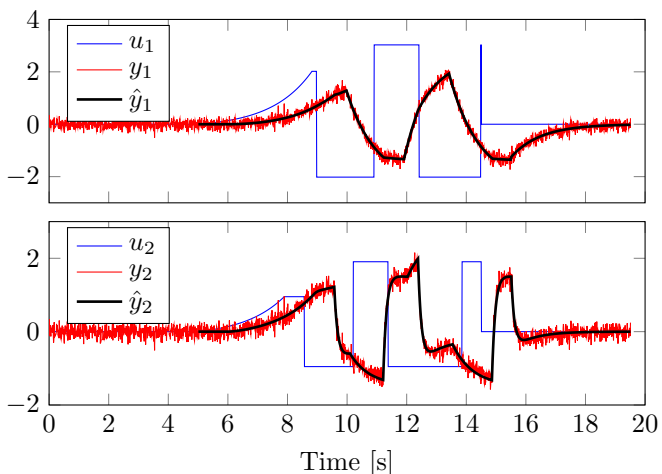| Process | $\hat{K}$ | $\hat{T}$ | $\hat{L}$ | $K$ | $T$ | $L$ |
|---|---|---|---|---|---|---|
| $G_{\epsilon=0}$ | 0.99 | 1.00 | 1.00 | 1 | 1 | 1 |
| | 0.01 | 0.99 | 1.15 | 0 | - | - |
| | 0.00 | 0.09 | 0.03 | 0 | - | - |
| | 1.00 | 0.10 | 1.00 | 1 | 0.1 | 1 |
| $G_{\epsilon=0.3}$ | 0.98 | 0.99 | 1.00 | 1 | 1 | 1 |
| | 0.32 | 1.02 | 1.02 | 0.3 | 1 | 1 |
| | 0.25 | 0.68 | 1.12 | 0.3 | 1 | 1 |
| | 1.02 | 0.11 | 1.00 | 1 | 0.1 | 1 |
| $G_{\epsilon=0.8}$ | 0.97 | 0.96 | 0.99 | 1 | 1 | 1 |
| | 0.85 | 1.17 | 1.00 | 0.8 | 1 | 1 |
| | 0.76 | 0.98 | 1.01 | 0.8 | 1 | 1 |
| | 1.03 | 0.13 | 0.99 | 1 | 0.1 | 1 |
| $G_{\mathrm{WB}}$ | 15.9 | 21.0 | 1.03 | 12.8 | 16.7 | 1 |
| | -18.2 | 20.7 | 3.00 | -18.9 | 21 | 3 |
| | 5.84 | 9.62 | 7.02 | 6.6 | 10.9 | 7 |
| | -20.2 | 15.1 | 2.99 | -19.4 | 14.4 | 3 |
| $G_{\mathrm{SOTD}}$ | 1.13 | 2.09 | 1.43 | 1 | {1,1} | 1 |
| | 0.32 | 2.17 | 1.33 | 0.3 | {1,1} | 1 |
| | 0.32 | 1.90 | 1.47 | 0.3 | {1,1} | 1 |
| | 1.02 | 0.17 | 1.04 | 1 | {0.1,0.1} | 1 |



Fig. 7. Experiment data and model fit for the intermediately coupled example $G_{\epsilon=0.3}$.
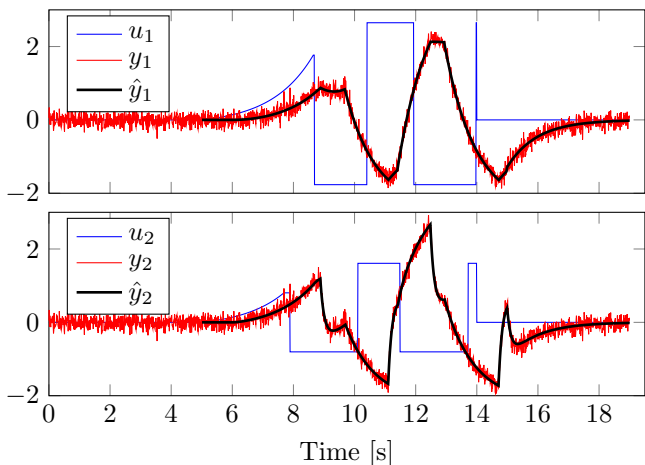


Fig. 8. Experiment data and model fit for the strongly coupled example $G_{\epsilon=0.8}$.
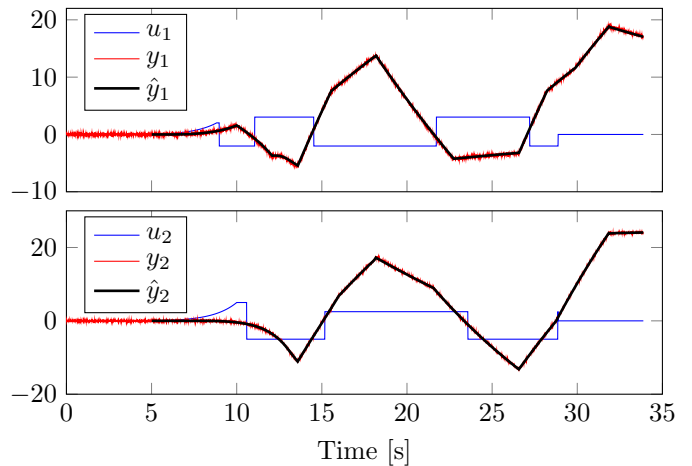


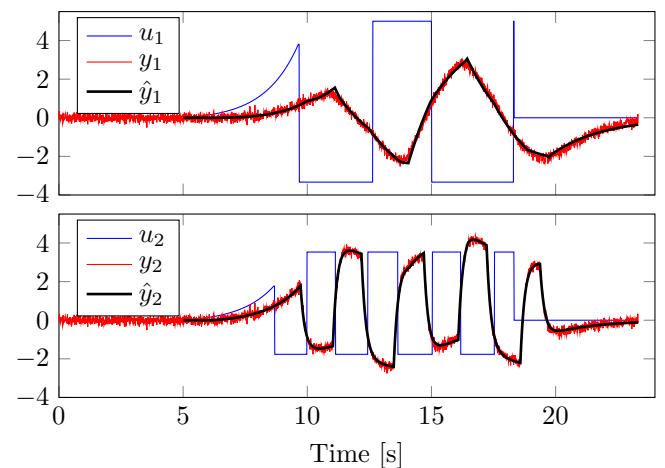Fig. 9. Experiment data and model fit for the Wood-Berry distillation column $G_{\mathrm{WB}}$.



Fig. 10. Experiment data and model fit for the second order time delayed process $G_{\mathrm{SOTD}}$.

## 4. DISCUSSION

The proposed method shows that good process models of a TITO system of FOTD subprocesses can be identified by one single decentralized relay experiment. There is no need to wait for convergence of the limit cycle oscillations and neither to fully understand the conditions for when the different limit cycles occur. The method is shown to work well for the investigated FOTD examples. The result for the SOTD system is also satisfying in the sense that it provides good FOTD approximations of the processes. However, better control performance can be obtained if a higher order model was estimated instead. As stated in Sec. 2.3 the method can be extended to estimate higher order models, and a generalization of the method to both higher order models as well as arbitrary MIMO systems to see how it scales, and if the excitation from one single experiment is sufficient, are interesting continuations of this work. Practical issues like parameter choices, experiment start-up and termination, and of course how the controllers should be tuned from the obtained models, are other aspects that should be explored further. Noise has been considered in this paper, but the method's sensitivity to other disturbances has not yet been investigated.
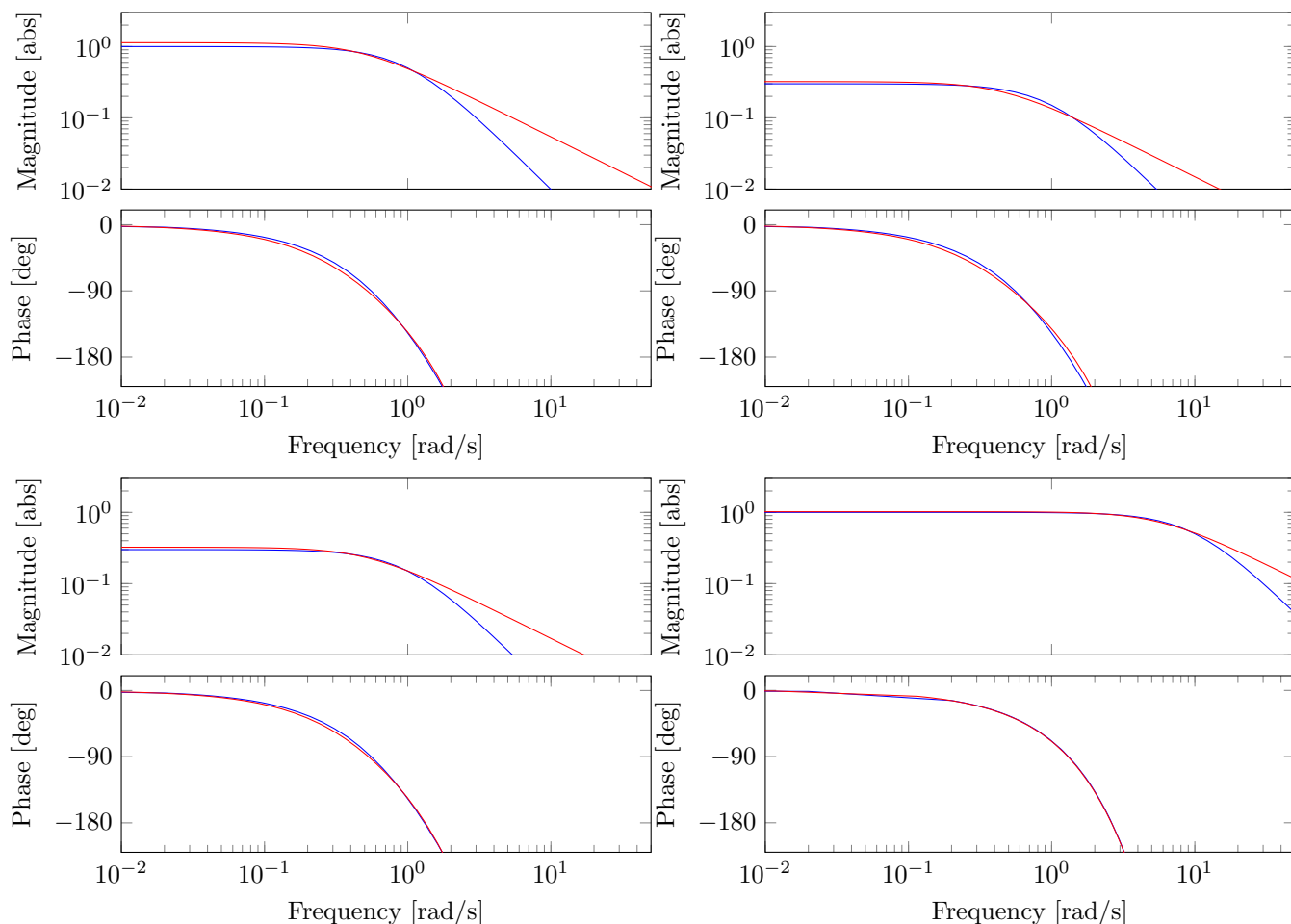
Fig. 11. Bode plots for the true SOTD TITO process in blue, and the estimated FOTD TITO model in red.

## REFERENCES

Åström, K.J. and Bohlin, T. (1965). *Numerical identification of linear dynamic systems from normal operating records*, 96–111. Springer, New York.

Åström, K.J. and Hägglund, T. (1984). Automatic tuning of simple regulators with specifications on phase and amplitude margins. *Automatica*, 20(5), 645–651.

Berner, J., Hägglund, T., and Åström, K.J. (2016a). Asymmetric relay autotuning–Practical features for industrial use. *Control Engineering Practice*, 54, 231–245.

Berner, J., Hägglund, T., and Åström, K.J. (2016b). Improved relay autotuning using normalized time delay. In *2016 American Control Conference (ACC)*, 1869–1875. IEEE.

Boyd, S., Hast, M., and Åström, K.J. (2015). MIMO PID tuning via iterated LMI restriction. *International Journal of Robust and Nonlinear Control*.

Campestrini, L., Barros, P.R., and Bazanella, A.S. (2006). Auto-tuning of PID controllers for MIMO processes by relay feedback. *IFAC Proceedings Vol.*, 39(2), 451–456.

Chidambaram, M. and Sathe, V. (2014). *Relay Autotuning for Identification and Control*. Cambridge University Press.

Liu, T., Wang, Q.G., and Huang, H.P. (2013). A tutorial review on process identification from step or relay feedback test. *Journal of Process Control*, 23(10), 1597–1623.

Loh, A. and Vasnani, V. (1994). Necessary conditions for limit cycles in multiloop relay systems. *IEE Proceedings-Control Theory and Applications*, 141(3), 163–168.

Luyben, W.L. (1987). Derivation of transfer functions for highly nonlinear distillation columns. *Ind. Eng. Chem. Res.*, 26(12), 2490–2495. doi:10.1021/ie00072a017.

Shen, S.H., Wu, J.S., and Yu, C.C. (1996). Use of biased-relay feedback for system identification. *AIChE J.*, 42(4), 1174–1180. doi:10.1002/aic.690420431.

Soltesz, K., Hägglund, T., and Åström, K.J. (2010). Transfer function parameter identification by modified relay feedback. In *American Control Conference*, 2164–2169. Baltimore, MD.

Wang, Q.G., Zou, B., Lee, T.H., and Bi, Q. (1997). Autotuning of multivariable PID controllers from decentralized relay feedback. *Automatica*, 33(3), 319–330.

Wood, R. and Berry, M. (1973). Terminal composition control of a binary distillation column. *Chemical Engineering Science*, 28(9), 1707–1717.

Yu, C.C. (2006). *Autotuning of PID controllers: A relay feedback approach*. Springer Science & Business Media.