Stochastic time series forecasting using time-delay reservoir computers: performance and universality

Lyudmila Grigoryeva[1], Julie Henriques[1,2], Laurent Larger[1], and Juan-Pablo Ortega[3]

[1]Université de Franche-Comté [2]Cegos Deployment [3]CNRS, Université de Franche-Comté

Context and objectives

The project [4] is devoted to a recently introduced machine learning paradigm called *Reservoir Computing (RC)*, also referred to in the literature as liquid state machines, echo state networks, or nonlinear transient computing. We study *time-delay reservoirs (TDRs)* that are constructed out of the sampling of the solution of a time-delay differential equation and show their good performance in the forecasting of the conditional covariances associated to multivariate discrete-time nonlinear stochastic processes of VEC-GARCH type as well as in the prediction of factual daily market realized volatilities computed with intraday quotes, using as training input daily log-return series of moderate size. We tackle some problems associated to the lack of task-universality for individually operating reservoirs and propose a solution based on the use of parallel arrays of time-delay reservoirs.

Stochastic nonlinear time series forecasting using TDRs

The VEC-GARCH family is the multivariate extension of the one-dimensional generalized autoregressive conditionally heteroscedastic (GARCH) models [3, 2]. The VEC-GARCH(1,1) model is determined by:

$$\begin{cases} \mathbf{z}_t = H_t^{1/2}\boldsymbol{\epsilon}_t \quad \text{with} \quad \{\boldsymbol{\epsilon}_t\} \sim \text{IIDN}(\mathbf{0}, \boldsymbol{I}_n), \\ \boldsymbol{h}_t = \mathbf{c} + A\boldsymbol{\eta}_{t-1} + B\boldsymbol{h}_{t-1}, \end{cases} \quad (1)$$

where $\{\mathbf{z}_t\}$ is n-dimensional conditionally heteroscedastic discrete-time process, $\{H_t\}$ is a conditional covariance matrix process of $\{\mathbf{z}_t\}$, $\boldsymbol{h}_t := \text{vech}(H_t)$, $\boldsymbol{\eta}_t := \text{vech}(\mathbf{z}_t\mathbf{z}_t^T)$, $\mathbf{c} \in b\mathbf{R}^N$, $N := n(n+1)/2$, and $A, B \in \mathbb{M}_N$. The model requires $N(2N+1)$ parameters for a complete specification.

Volatility forecasting. The main use of the VEC(1,1) model (1) is the forecasting of the volatility of financial time series: $\{\mathbf{z}_t\}$ are considered as market log-returns and $\{H_t\}$ are the associated conditional covariance matrices. The optimal h-step ahead forecast $\widehat{\mathbf{h}_{T+h}}$ for the covariance matrix $\mathbf{h}_{T+h}$ knowing the log-returns $\{\mathbf{z}_0, \ldots, \mathbf{z}_T\}$ is given by the conditional expectation with respect to that information set:

$$\widehat{\mathbf{h}_{T+h}} := E\left[\mathbf{h}_{T+h} \mid \mathcal{F}_T\right].$$

For the VEC-GARCH(1,1) it can be explicitly computed via the recursion on the horizon h:

$$\begin{aligned} \widehat{\mathbf{h}_{T+1}} &= \mathbf{h}_{T+1} = \mathbf{c} + A\boldsymbol{\eta}_T + B\boldsymbol{h}_T, \\ \widehat{\mathbf{h}_{T+2}} &= \mathbf{c} + (A+B)\widehat{\boldsymbol{h}_{T+1}}, \\ &\vdots \\ \widehat{\mathbf{h}_{T+h}} &= \mathbf{c} + (A+B)\widehat{\mathbf{h}_{T+h-1}}. \end{aligned} \quad (2)$$

The functional dependence of the forecast $\widehat{\mathbf{h}_{T+h}}$ on the elements $\{\mathbf{z}_0, \ldots, \mathbf{z}_T\}$ that generate the information set $\mathcal{F}_T$ is quadratic.

The objective is to prove empirically that TDRs are capable of forecasting performances comparable to those attained using the parametric Box-Jenkins approach when the nonlinear VEC-GARCH volatility models are taken as data generating process (DGP).

TDR architecture

We work with a TDR with the same nonlinear kernel as in [1]

$$f(x(t-\tau), u(t), \eta, \gamma, p) = \frac{\eta\,(x(t-\tau) + \gamma u(t))}{1 + (x(t-\tau) + \gamma u(t))^p}, \quad (3)$$

where γ, η, and p are real valued parameters, and τ is the time delay. Reservoirs contain N neurons with separation $\theta := \tau/N$. As a teaching signal we use $\{\mathbf{h}_{1+h}, \ldots, \mathbf{h}_{T_{\text{train}}}\}$ generated by the model together with the time series values $\{\mathbf{z}_1, \ldots, \mathbf{z}_{T_{\text{train}}-h}\}$, that is, $\mathbf{y}(k) := \mathbf{h}_{k+h} \in \mathbb{R}^n$, $k \in \{1, \ldots, T_{\text{train}} - h\}$ and the trained output layer is given by the solution of the ridge (or Tikhonov) regression

$$W_{\text{out}} := (XX^T + \lambda\mathbb{I}_N)^{-1}X\mathbf{y}, \quad (4)$$

where $X \in \text{Mat}(N, T_{train})$ is the reservoir output given by $X_{i,j} := x_i(j)$ and $\mathbf{y} \in \text{Mat}(T_{train}, n)$ is the teaching matrix containing the vectors $\mathbf{y}(k)$, $k \in \{1, \ldots, T_{train}\}$, organized by rows, λ is the strength of the ridge penalty introduced in order to regularize the regression problem hence helping to avoid overfitting.

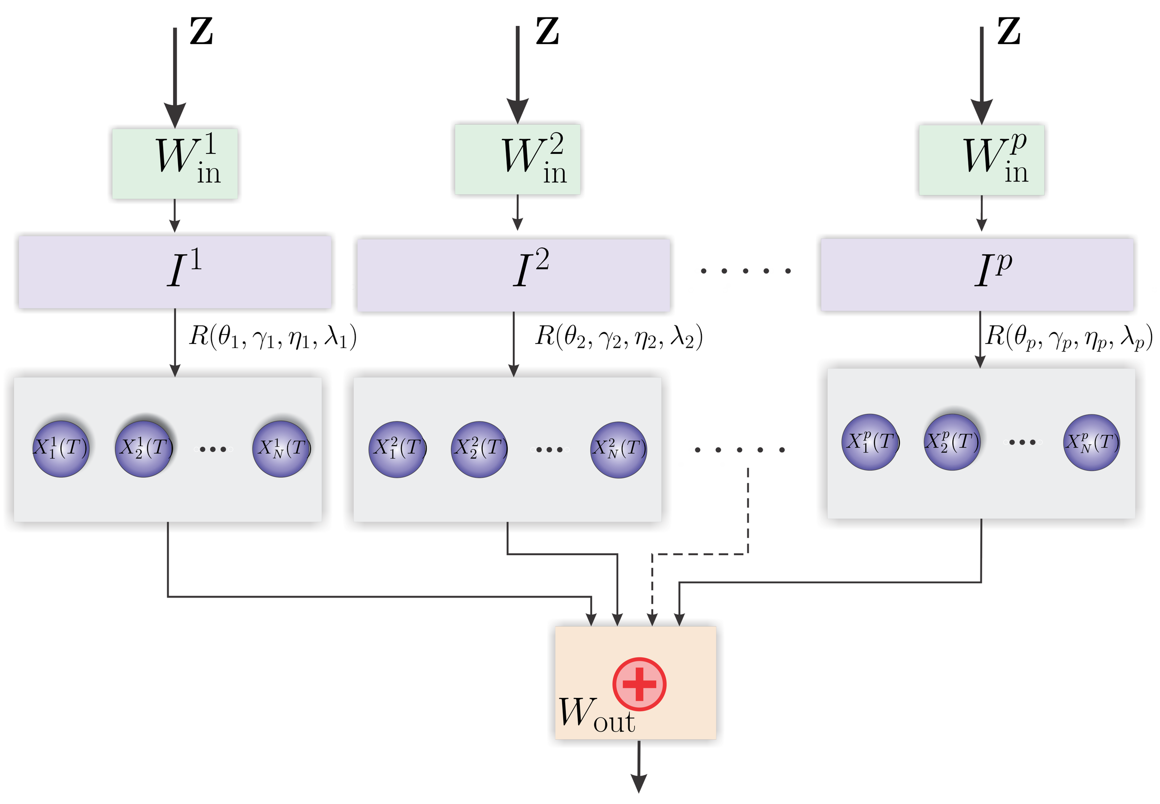

Figure 1. Parallel array of p TDRs.

Reservoir configurations under study

The different devices whose performances are compared have the following architectures:

1. *Individually operating* TDR of 400 neurons and grid optimized parameters;
2. *Individually operating* TDR of 400 neurons and random optimized parameters;
3. *Parallel array* of 40 TDRs of 10 neurons each with random optimized parameters (Figure 1);
4. *Parallel array* of 80 TDRs of 5 neurons each with random optimized parameters (Figure 1).

Empirical results with synthetic data

1. All the TDRs considered are capable of competitively accomplishing forecasting tasks for stochastic nonlinear processes without the need to solve the sophisticated model selection and estimation problems that arise in the parametric approach;
2. The good performance in the forecasting task is mostly due to the presence of the time-delay reservoir;
3. The kernel parameters for the individually operating TDR need to be optimized in order to achieve adequate predictions;
4. Advantages of using the parallel pools of TDRs: possibility of random parameter optimization and limitation of the computational effort (Figure 2), better performance for smaller training sample sizes (Figure 3), and improved universality with respect to changes in the forecasting horizon (Figure 4) and in the model specification (Figure 5).

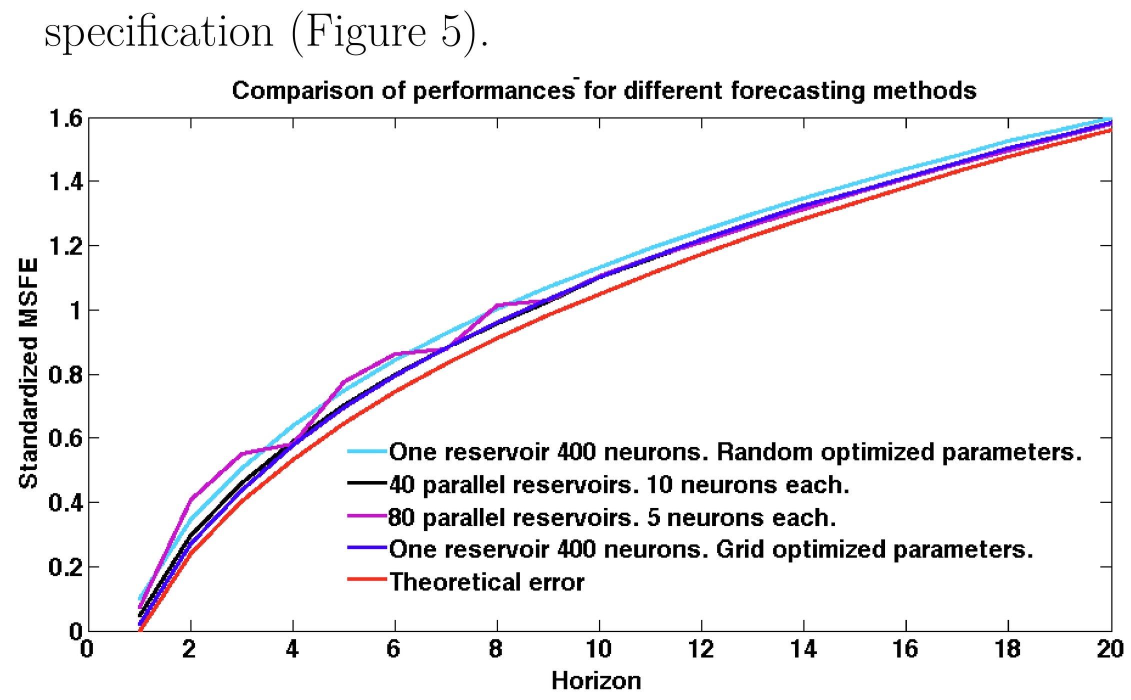

Figure 2. Forecasting performance in the three dimensional VEC-GARCH volatility forecasting task achieved with the different RC based methods in comparison with the theoretical error as a function of the forecasting horizon.

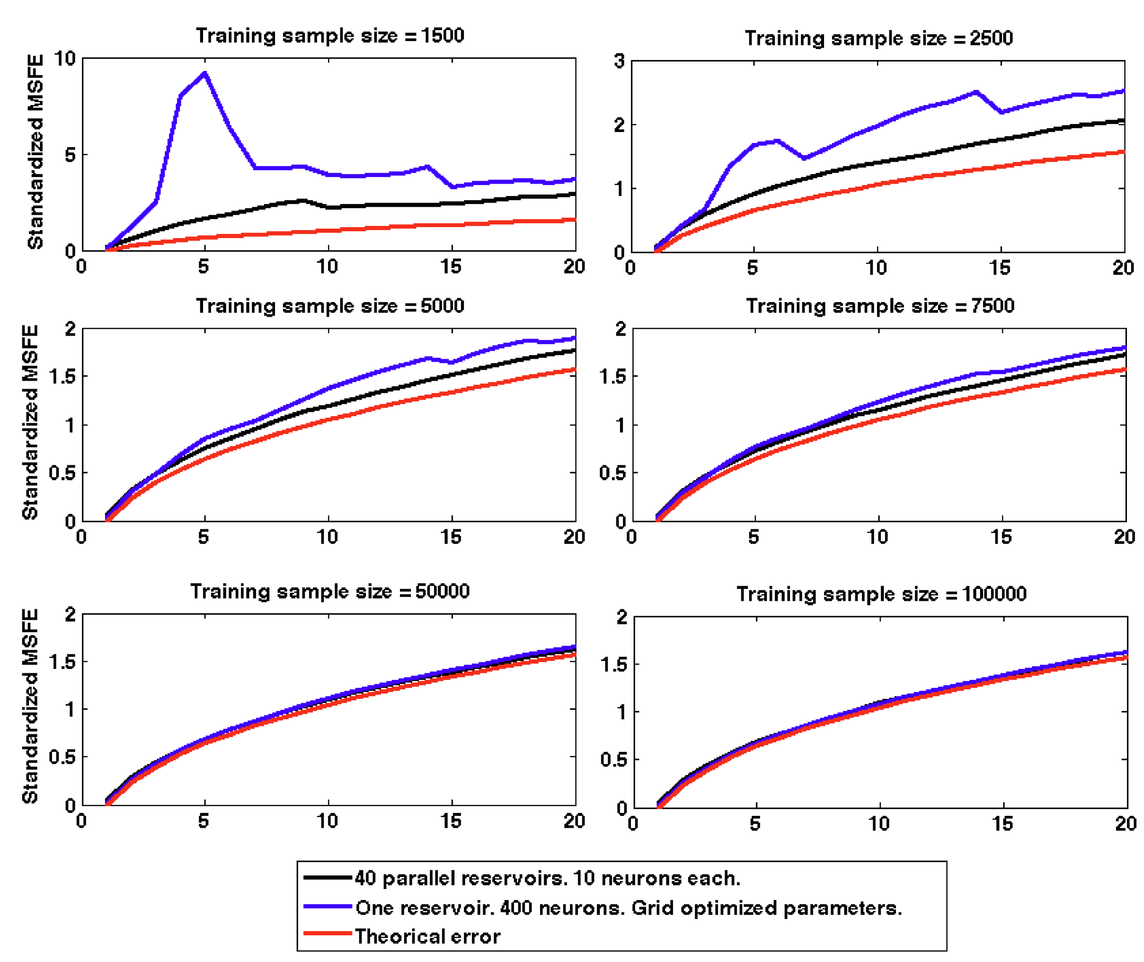

Figure 3. Comparison of the sMSFE committed for different training sample sizes by a single grid optimized TDR with 400 neurons and by the parallel pools of TDRs.

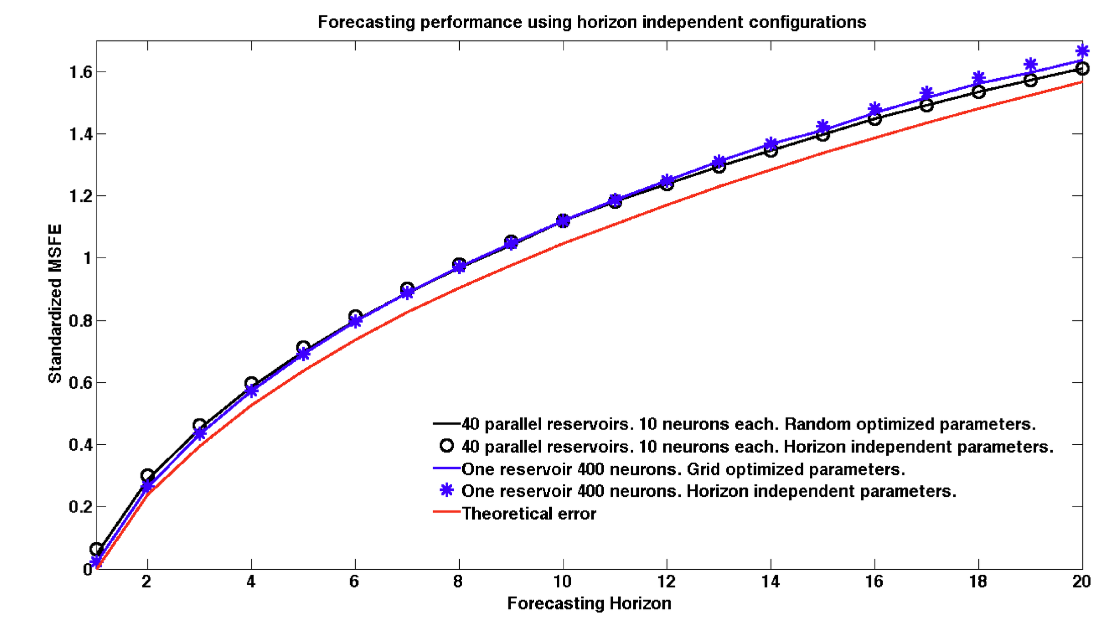

Figure 4. Comparison of the forecasting performances obtained by using horizon adapted parameter configurations and constant parameters.

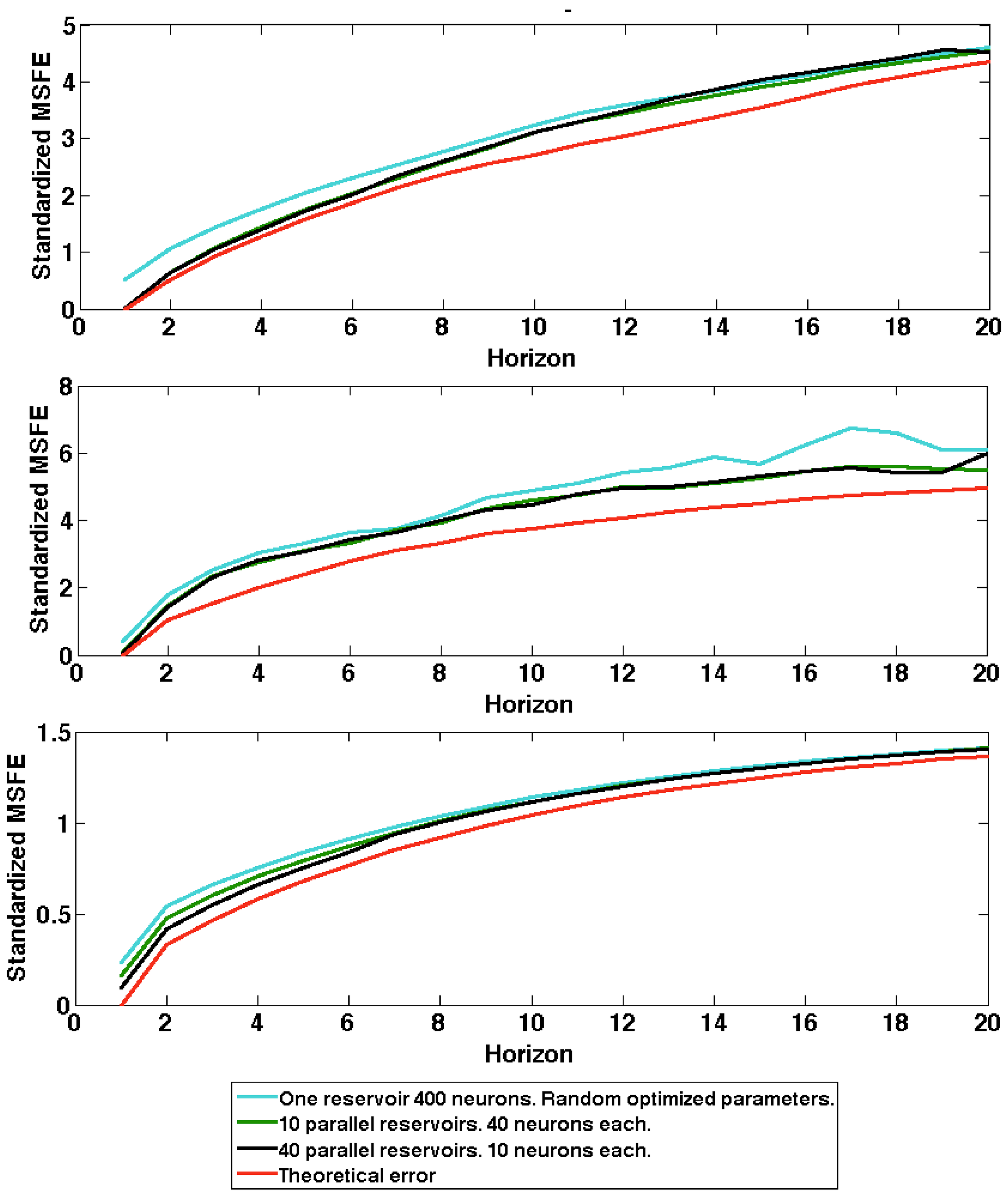

Figure 5. Forecasting performance for three misspecified models.

Empirical results with real market data

The dataset consists of 2483 daily log-returns for the NYSE quoted assets with Yahoo tickers AAPL, ABT, AXP, BA, BAC, BMY, BP, C, and CAT (Jan 6th, 1999–Dec 31st, 2008). The realized covariance matrices $\{H_t\}$ are constructed using six minutes sampled intraday data. The first 2000 observations are used for training/estimation purposes and the remaining part is used for an out-of-sample test.

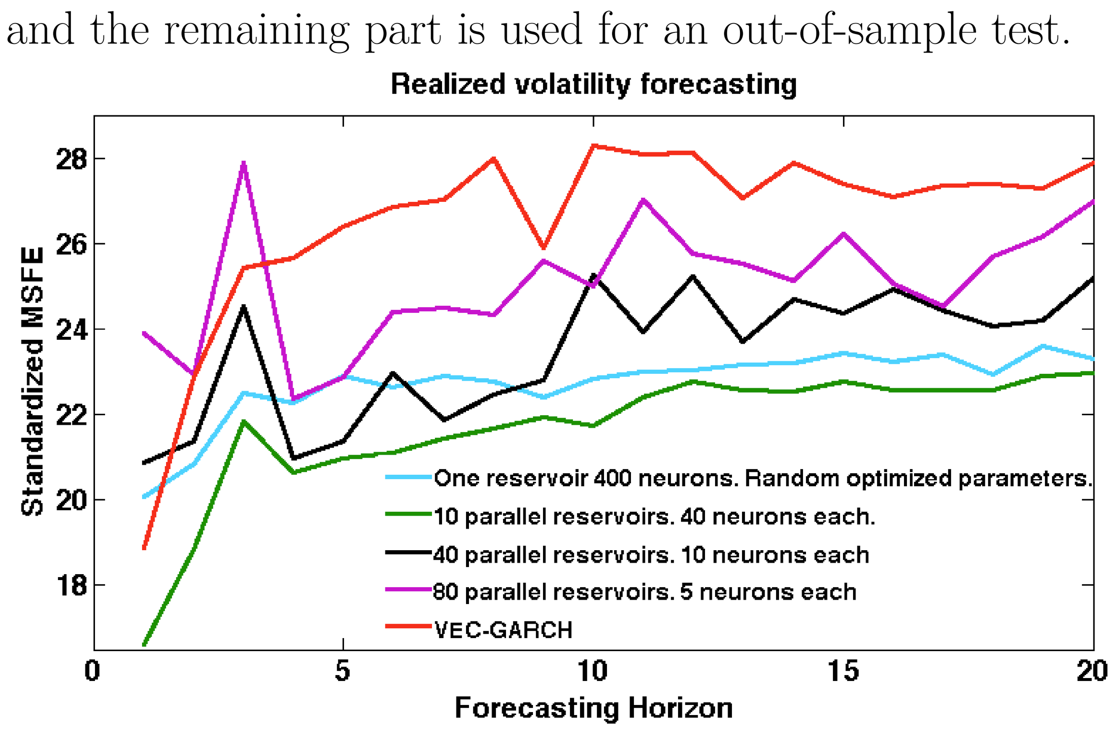

Figure 6. The sMSFE reported is an average over the sMSFEs associated to the realized volatility forecasting task for the 84 different three dimensional combinations of the nine assets considered. For each asset combination, forecasting with the VEC(1,1) approach is carried out with a different model estimated via MLE using the historical evolution of that particular combination. In the case of RC, for a given horizon, a single parameter set is used for all the combinations, chosen by minimizing the training error.

References

[1] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer. Information processing using a single dynamical node as complex system. ***Nature communications***, 2:468, January 2011.

[2] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. ***Journal of Econometrics***, 31(3):307–327, 1986.

[3] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. ***Econometrica***, 50(4):987–1007, 1982.

[4] Lyudmila Grigoryeva, Julie Henriques, Laurent Larger, and Juan-Pablo Ortega. Stochastic time series forecasting using time-delay reservoir computers: performance and universality. ***http://dx.doi.org/10.2139/ssrn.2350331***, 2013.

Acknowledgements

We acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493), the European project PHOCUS (FP7 Grant No. 240763), the Labex ACTION program (Contract No. ANR-11-LABX-01-01), and Deployment S.L. LG acknowledges financial support from the Faculty for the Future Program of the Schlumberger Foundation.

LabeX Action Integrated smart systems