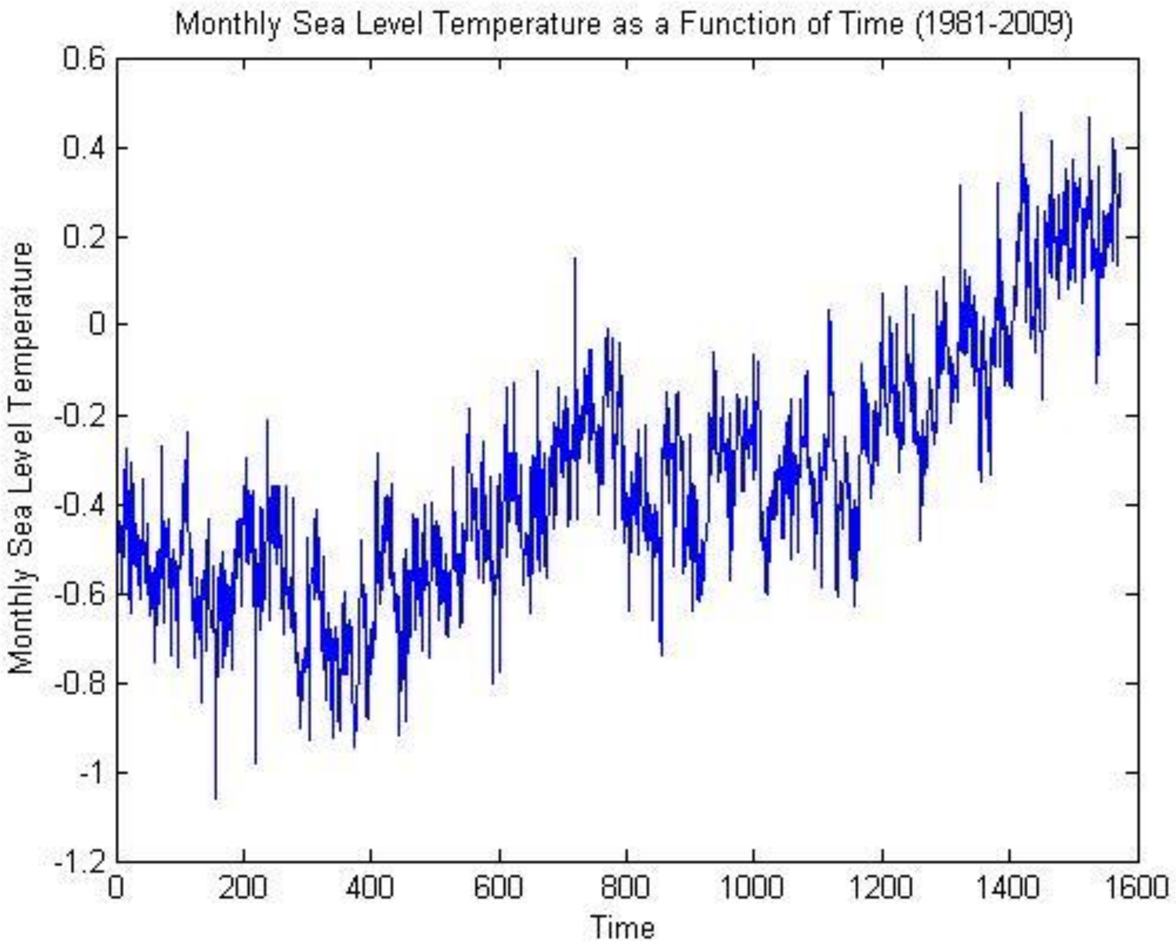
Time Series Analysis: an online learning approach

Oren Anava
Elad Hazan
Shie Mannor
Ohad Shamir

Technion
Israel Institute of Technology

COLT 2013

Time Series



Successive measurements, unif spaced

Analysis and Forecasting

- Rich history in statistics/economics:
  - Yule 1920's
  - Whittle 1951
  - Box & Jenkins, 1970's (max-likelihood estimation)
  - Engle 1982
- Numerous models:
  - AR, MA, ARMA, ARIMA, ARCH, GARCH, …
- Focus on parameter estimation (Yule-Walker equations)
- Mean Square Error criterion only
- Assume uncorrelated i.i.d Gaussian noise (using the MSE criterion, w.l.o.g)

ARMA model

- AR(k) model:

$$X_t = \sum_{i=1}^{k} \alpha_i X_{t-1} + \varepsilon_t$$

- ARMA(k,q) model:

$$X_t = \sum_{i=1}^{k} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t$$

Predicting Time Series

Game protocol:

- Adversary fixes (α,β)
- At iteration t, adversary samples ε_t, generates

$$X_t(\alpha,\beta) = \sum_{i=1}^{k} \alpha_i X_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \varepsilon_t$$

- Online player predicts Y_t, suffers loss:

$$\ell(X_t, Y_t)$$

- Goal: minimize (average) regret

$$\frac{1}{T}\sum_{t=1}^{T}\ell(X_t,Y_t) - \frac{1}{T}\min_{\alpha,\beta} E\sum_{t=1}^{T}\ell(X_t, X_t(\alpha,\beta))$$

Our results

For convex loss functions, ALG1, Regret = $O(\frac{\log T}{\sqrt{T}})$
Linear running time, almost tight.

Exp-concave loss functions, ALG2, Regret = $O(\frac{\log^2 T}{T})$
Quadratic running time, almost tight

Weak assumptions on noise…
Assume:

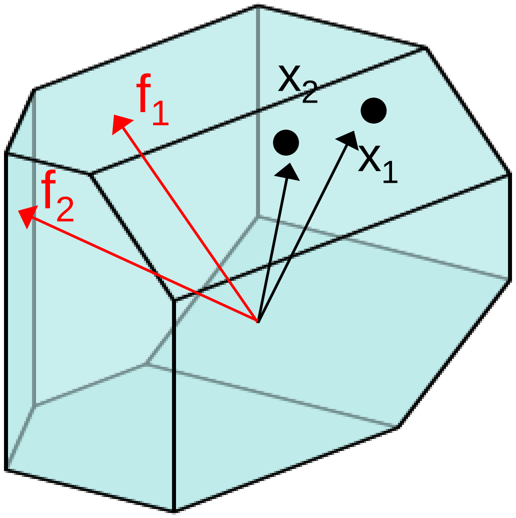
- $|\alpha_i| < c < \infty$ $\sum_{j=1}^{q} |\beta_j| < 1$
- Loss functions are Lipschitz and bounded

Our results vs. Prev. results

	Previous Work	Our Result
Square loss (MSE)	Ding et al: $Regret = O(\frac{\log^{2+\varepsilon} T}{T})$	$Regret = O(\frac{\log^2 T}{T})$
General Loss	NONE	$O(\frac{1}{\sqrt{T}})$

- weaker noise assumptions.
- General loss functions - reduction to Gaussians fails.

Online convex optimization



Incurred loss

$f_1(x_1)$

$f_2(x_2)$

$f_T(x_T)$

Convex (linear) bounded cost functions

Total loss = $\sum_t f_t(x_t)$

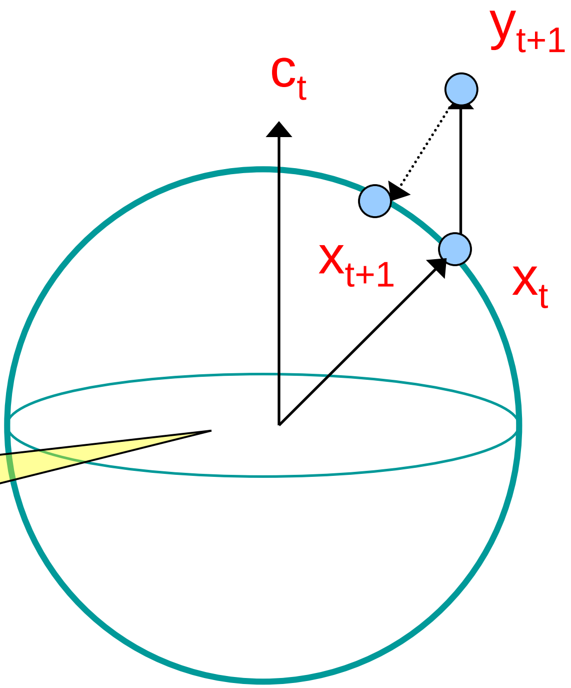
Regret = $\sum_t f_t(x_t) - \min_{x^*} \sum_t f_t(x^*)$

ARMA: x_t are (α_t, β_t)

Sublinear regret -> Risk bounds

Online gradient descent

The algorithm: move in the direction of the vector c_t (gradient of the current cost function)



$y_{t+1} = x_t - \eta\, c_t$

and project back to the convex set

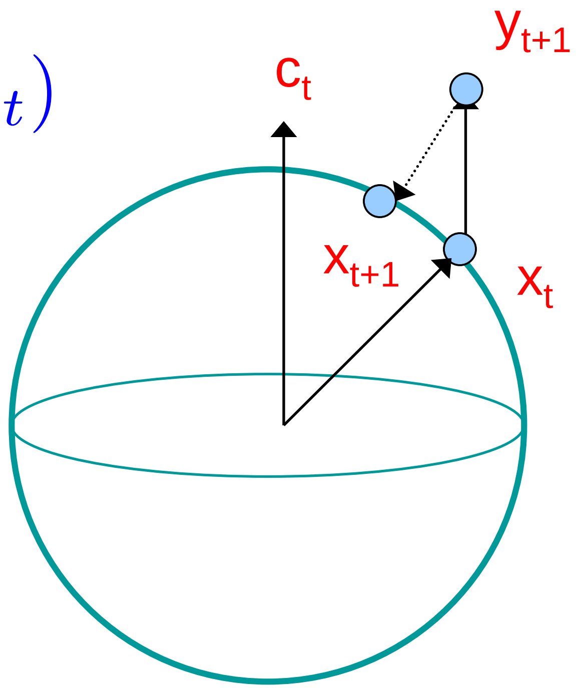
Thm [Zinkevich]: if $\eta = 1/\sqrt{t}$ then this alg attains worst case regret of

$\sum_t f_t(x_t) - \sum_t f_t(x^*) = O(\sqrt{T})$

Online mirrored descent

$$\nabla R(y_{t+1}) \leftarrow \nabla R(x_t) - \eta \nabla f_t(x_t)$$

$$x_{t+1} \leftarrow \arg\min_{x \in \mathcal{K}} B_R(x, y_{t+1})$$



Other algs: FTRL, Online Newton Step,FPL,…

Using OCO?

Learning model parameters?

Loss functions:

$$f_t(\alpha,\beta) \equiv \ell(X_t, \sum_{i=1}^{k} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j})$$

1. ε terms unknown
2. Function is not convex

Estimating the noise

Recursively define:

$$\tilde{X}_t^{\infty}(\alpha,\beta) \equiv \sum_{i=1}^{k} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j (X_{t-j} - \tilde{X}_{t-i}^{\infty}(\alpha,\beta))$$

Can define new loss functions based upon these estimates.

$$f_t^{\infty}(\alpha,\beta) = \ell(X_t, \sum_{i=1}^{t} c_i(\alpha,\beta) X_{t-i})$$

1. “infinite memory” – new loss functions depend on entire history.
2. Still not convex

Non-Proper learning

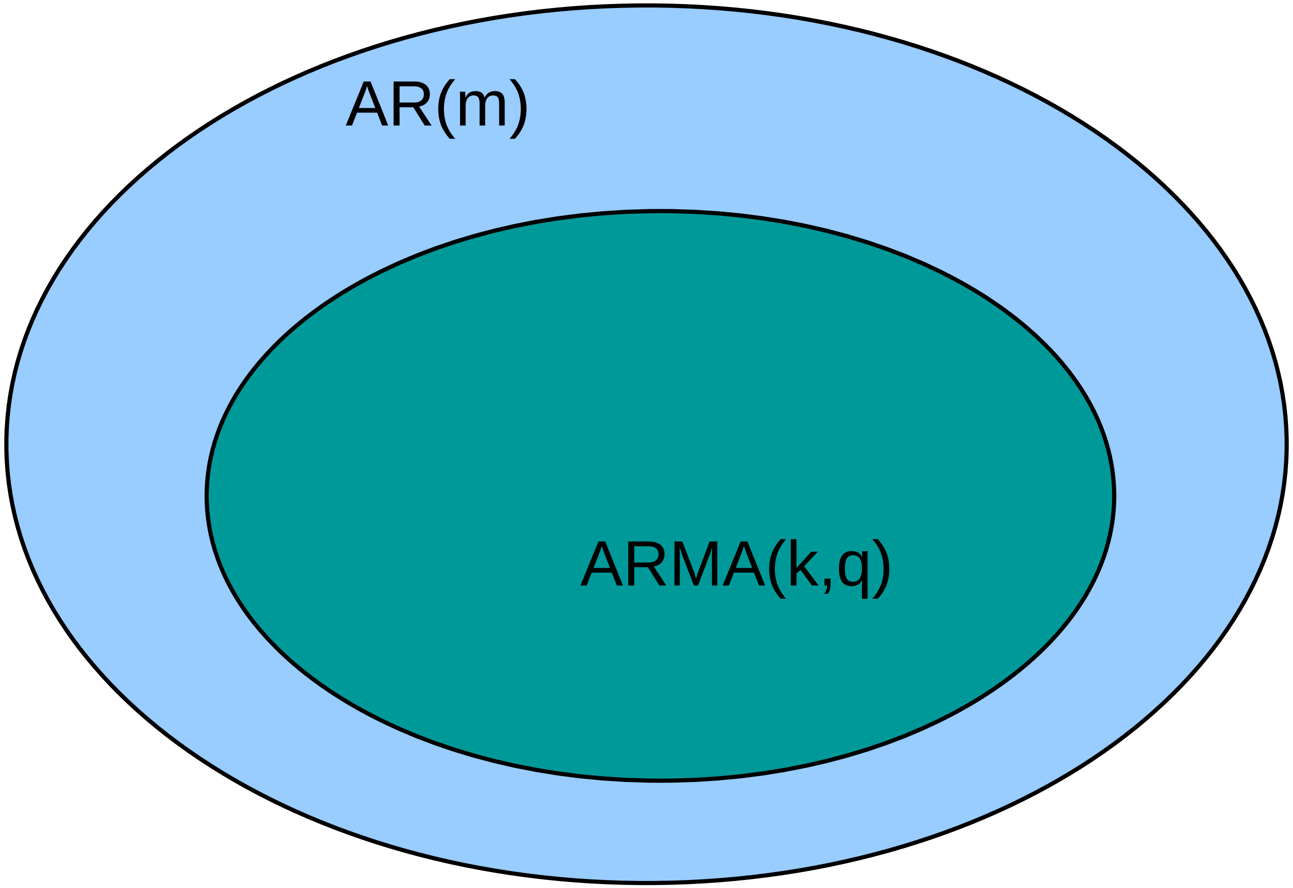
Predict with Y_t which is outside the model! Convexify by non-proper learning:

$$f_t^{\infty}(\gamma) = \ell(X_t, \sum_{i=1}^{t} \gamma_i X_{t-i})$$

Bounded memory:

We take: $$Y_t = \sum_{i=1}^{m} \gamma_i X_{t-i}$$

Main observation



$$ARMA(k, q) \subseteq AR(k + q \log T)$$

AR modeled by online convex optimization

$$f_t^m(\gamma) = \ell(X_t, \sum_{i=1}^{m} \gamma_i X_{t-i})$$

Are convex, learnable using OGD/ONS, with optimal rates.

$ARMA(k,q) \subseteq AR(k + q \log T)$ - follows by assumption on geometric decrease in beta, and structure of recursive noise estimation.

The algorithms

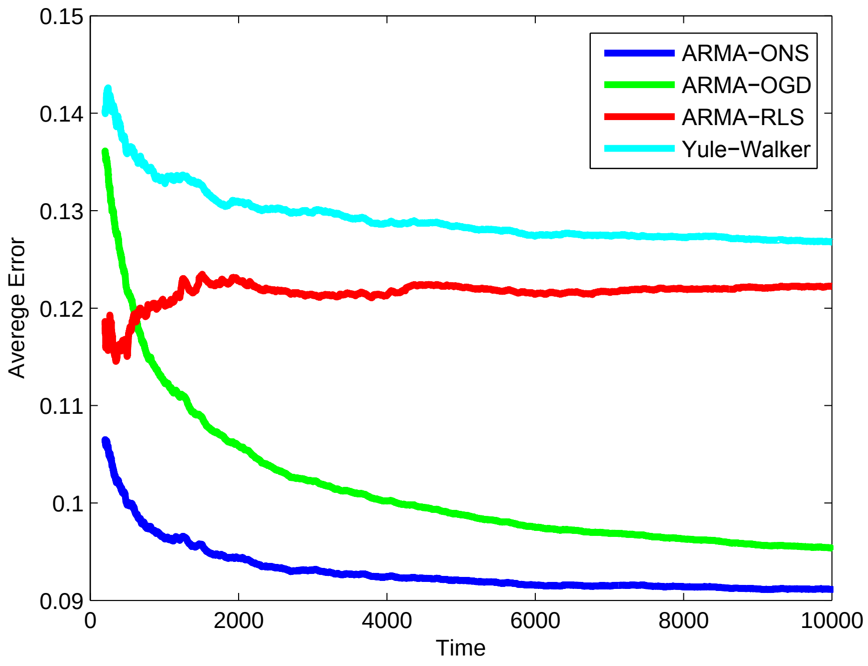
$$f_t^m(\gamma) = \ell(X_t, \sum_{i=1}^{m} \gamma_i X_{t-i})$$

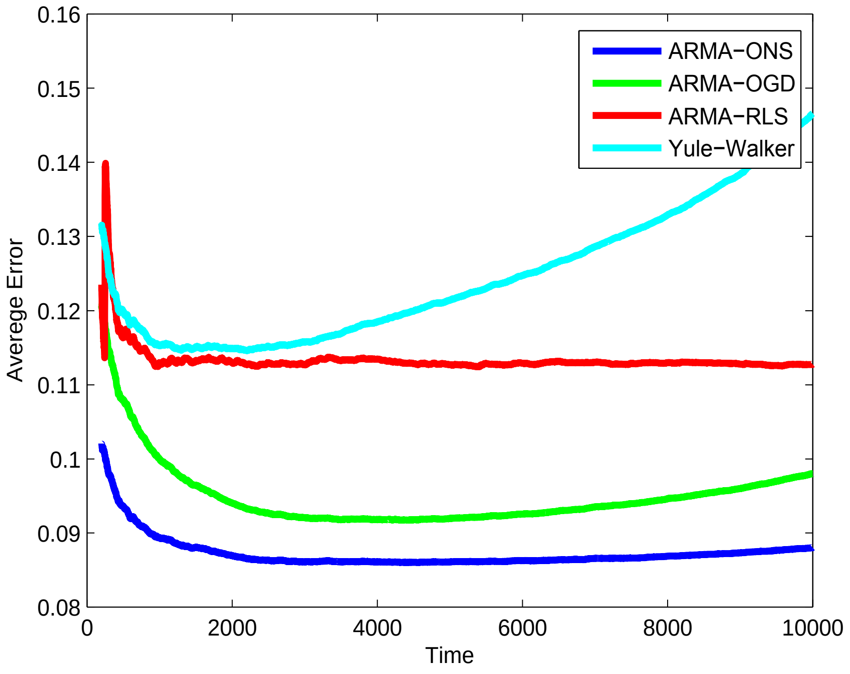
Apply Online Gradient Descent / Online Newton Step on the above…

For general convex loss (i.e. l1 or hinge), root(T) regret via:

$$\gamma_{t+1} \leftarrow \prod_{\mathcal{K}} [\gamma_t - \eta \nabla f_t(\gamma_t)] \qquad Y_t = \sum_{i=1}^{m} \gamma_i X_{t-i}$$

Experiments - generated data

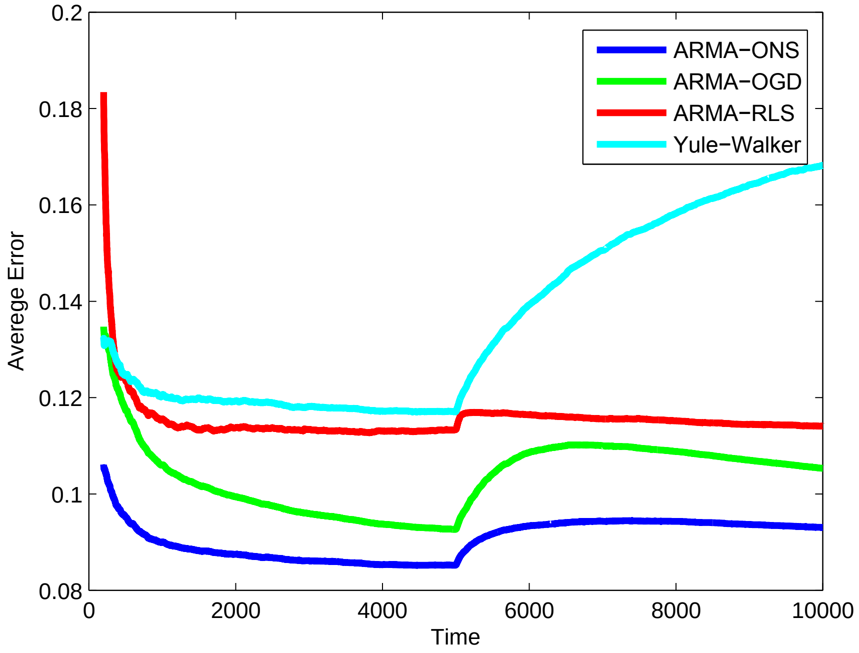


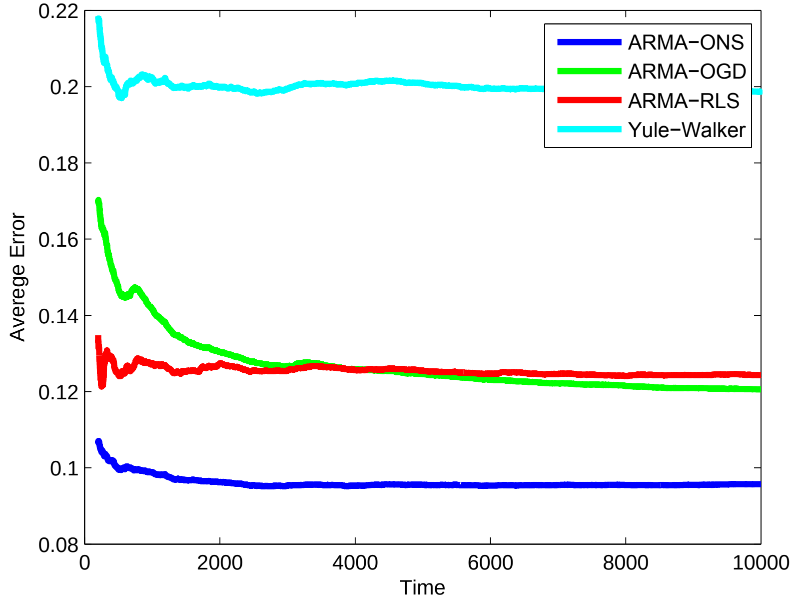


Sanity check –
Gaussian noise

Slowly changing
coefficients

Simulated data

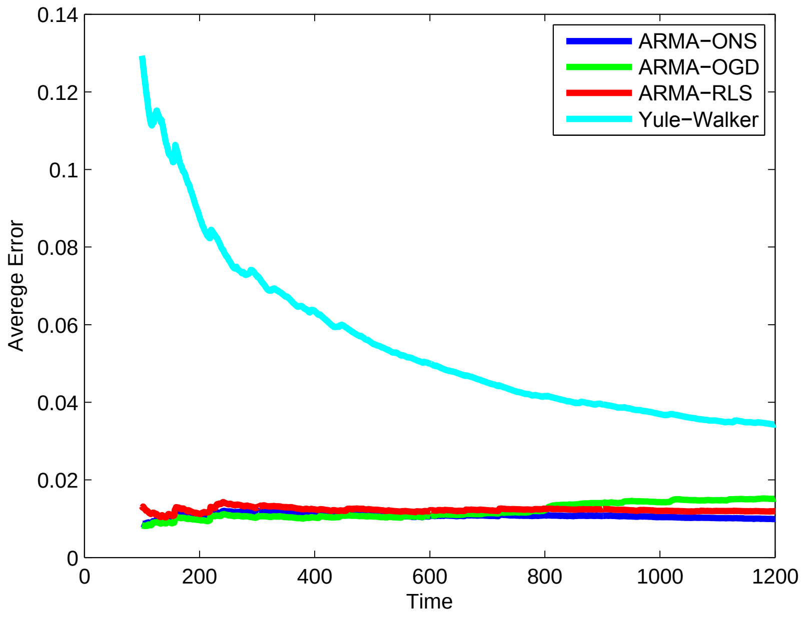


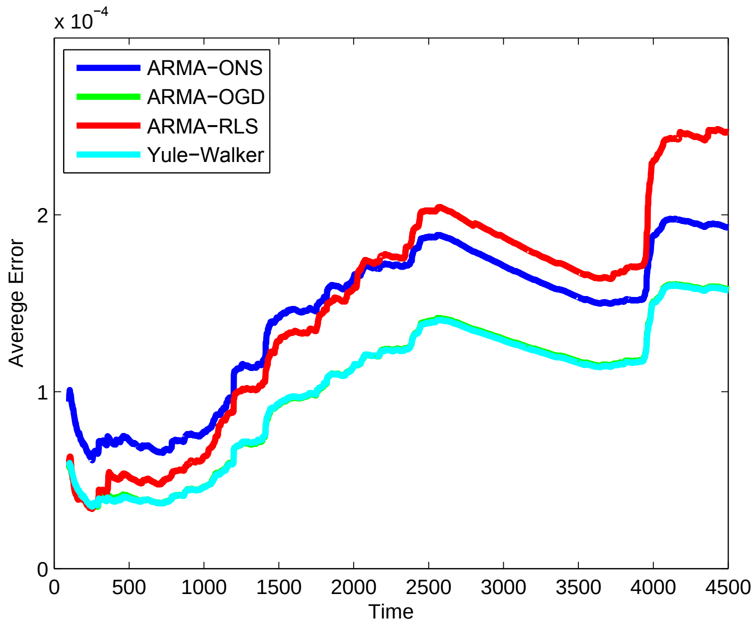


Abrupt change

Correlated noise

Real data





Temperature forcast

S&P 500

More research / open questions

- **New approach to time series analysis & forecasting –**
  Predict as best ARMA model under weak assumptions, efficiently
- **Promising empirical results, esp. second order algorithms**
- **Can we relax the assumptions on noise coefficients?**
- **Generalization to adversarial noise?**
- **Tackle ARCH/GARCH?**

Thank you!