Time series analysis for Financial Data

A S Vasudeva Murthy

TIFR-CAM, Bangalore

Saksham Bassi & Atharava Gomekar

Pune Eng. College.

Motivation

1. Machine Learning for time series
2. All India daily rainfall, onion prices, stock prices, astronomical data

Neural Network

Revived from 1990s due to large data and powerful computers.

Given $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$. Find x such that

$$Ax = b.$$

$$\text{If } det(A) \neq 0 \text{ then } x = A^{-1}b.$$

Now let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ for $m \neq n$. Find $x \in \mathbb{R}^n$ such that

$$Ax = b.$$

Minimize $J : \mathbb{R}^n \to \mathbb{R}$

$$J(x) = \|Ax - b\|_2$$

where $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots x_n^2}$ giving the solution (least squares)

$$x = (A^t A)^{-1} A^t b$$

obtained by using $\nabla J = 0$.
In general given a $f(x)$ we can approximate it by

$$F(x) = c_1\phi_1(x) + \cdots c_n\phi_n(x)$$

for a set of known $\{\phi_i(x)\}_{i=1}^n$ by minimizing $(\nabla J = 0)$

$$J(c) = \|f(x) - F(x)\|_2$$

where $\|f\|_2 = (\int f^2(x)dx)^{\frac{1}{2}}$.

Given

$$L \text{ layers with } n_j \text{ neurons at } j^{th} \text{ layer}$$

$$\text{Weight matrices } W^l \in \mathbb{R}^{n_l \times n_{l-1}}$$

$$\text{Baises } b^l \in \mathbb{R}^{n_l}$$

Then the neural network is defined by

$$y^1 = x$$

$$y^l(x) = \sigma(W^l y^{l-1}(x) + b^l), \quad l = 2, \cdots L$$

Here σ is the sigmoidal function

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

with the convention

$$\sigma(z_1, z_2 \cdots z_n) = (\ \sigma(z_1), \sigma(z_2) \cdots \sigma(z_n)\ ).$$

By back propagation we mean find W, b for a given data $\hat{y}_i$ observed at x_i for $i = 1, \cdots N$ that minimizes

$$J(W, b) = \frac{1}{N} \sum_{i=1}^{N} \|\{\hat{y}^i - y^l(x_i)\|$$

$$\nabla J \quad w.r.t. \quad W?$$

We recall specific cases. For $a, x \in R^{n \times 1}$

$$J(x) = a^t x, \quad \frac{\partial J}{\partial x} = a^t$$

For $A \in R^{n \times n}$

$$J(x) = x^t A x, \quad \frac{\partial J}{\partial x} = 2Ax$$

Let $X \in R^{m \times n}$

$$J(X) = a^t X b, \quad \frac{\partial J}{\partial X} = ab^t$$

If $X \in R^{n \times n}$

$$J(X) = a^t X a, \quad \frac{\partial J}{\partial X} = 2aa^t - D_{aa^t}$$

where D_{aa^t} is the diagonal of the matrix aa^t.

This is because for $i \neq j$

$$\partial J / \partial X_{ij} = a_i a_j + a_j a_i$$

and

$$\partial J / \partial X_{ii} = a_i^2.$$

If each $X_{pq} = f_{pq}(Y)$ where Y is another matrix then

$$\frac{\partial X}{\partial X_{pq}} = \frac{\partial f_{pq}(Y)}{\partial Y_{pq}}$$

and this will be a huge matrix and a pain to compute.

Math. Control Signals Systems (1989) 2: 303–314

Mathematics of Control, Signals, and Systems

Approximation by Superpositions of a Sigmoidal Function*

G. Cybenko†

Definition. We say that σ is *discriminatory* if for a measure $\mu \in M(I_n)$

$$\int_{I_n} \sigma(y^{\mathrm{T}}x + \theta)\, d\mu(x) = 0$$

for all $y \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$ implies that $\mu = 0$.

Theorem 1. *Let σ be any continuous discriminatory function. Then finite sums of the form*

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^{\mathrm{T}}x + \theta_j) \tag{2}$$

are dense in $C(I_n)$. In other words, given any $f \in C(I_n)$ and $\varepsilon > 0$, there is a sum, $G(x)$, of the above form, for which

$$|G(x) - f(x)| < \varepsilon \qquad \text{for all} \quad x \in I_n.$$

2 The proposed reconstruction

We propose an ANN for the process of reconstructing the time series from the matrix containing its statistics. With the given time series T = $\{t_1, t_2, \ldots, t_n\}$, we convert it into a matrix Y such that each set of 7 values of the time series T corresponds to one row of the matrix. We then calculate 8 statistics for each row of the matrix Y and compute another matrix $\Sigma(Y)$, where both rows of the said matrices have one to one mapping. We use the matrix $\Sigma(Y)$ as input to ANN and reconstruct the matrix Y. We denote, $\{\mathbf{y_1}, \ldots, \mathbf{y_n}\}$ as vectors for each row of the matrix Y, and $\{\mu, \gamma, \kappa, \ldots, S_t\}$ as the choices of the statistics and are described in Table 2.

$$Y = \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_n} \end{bmatrix} \rightarrow \Sigma(Y) = \begin{bmatrix} \mu(\mathbf{y_1}) & \gamma(\mathbf{y_1}) & \kappa(\mathbf{y_1}) & \ldots & S_t(\mathbf{y_1}) \\ \mu(\mathbf{y_2}) & \gamma(\mathbf{y_2}) & \kappa(\mathbf{y_2}) & \ldots & S_t(\mathbf{y_2}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu(\mathbf{y_n}) & \gamma(\mathbf{y_n}) & \kappa(\mathbf{y_n}) & \ldots & S_t(\mathbf{y_n}) \end{bmatrix}$$

We make the reconstruction robust by minimizing the mean squared error at each subsequent layer. The error function is defined as

$$\mathcal{L}(Y, Y') = ||Y - \sigma(W\Sigma(Y) + b)||^2$$

$$\mathcal{L}(W, b)$$

Statistic	Meaning
Mean (μ)	Central value of a discrete set of numbers
Skewness (γ)	Measure of asymmetry of a distribution of a random variable around its mean
Kurtosis (κ)	Measure of the peakness of data
Variance (σ^2)	Measures of how far a set of (random) numbers are spread out from their average value
Beta or Slope (β)	The slope of a line fitted on the sample
Standard Deviation (σ)	Measure to quantify the amount of variation or dispersion of a set of data values
Entropy (e)	Measure of regularity or unpredictability
Exponential Moving Average (S_t)	Series of averages of different subsets of the full data set

1. **Mean**: For n elements in one row, the mean is

$$\bar{x} = 1/N * \sum_{n=1}^{N} x_n .$$

2. **Skewness**: The sample skewness is computed as the Fisher-Pearson coefficient of skewness, i.e. $g_1 = m_3 / m_2^{3/2}$ where $1/N \sum_{n=1}^{N} (x[n] - \bar{x})^i$ is the biased i^th sample central moment and $\bar{x}$ is the sample mean.

3. **Kurtosis**: Fourth central moment divided by the square of the variance, i.e. $(1/N \sum_{n=1}^{N} (x[n] - \bar{x})^4) / \sigma^4$

4. **Variance:** The formula for variance is $\sigma^2 = \frac{\sum_{n=1}^{N} (x_n - \bar{x})^2}{N}$

5. **Beta**: Slope of the linear regression line is given by

$$\frac{N(\Sigma xy) - (\Sigma x)(\Sigma y)}{N(\Sigma x^2) - (\Sigma x)^2}$$

6. **Entropy**: Algorithm to find entropy:

```
a.value, counts = unique(data_row)
```

`counts` have the count of all the unique elements.

```
b.S = -sum(counts * log(counts))
```

`S` is the calculated entropy

7. **Exponential Moving Average:** $EMA = x_t * k + EMA_y * (1 - k)$

t is today, y is yesterday, $k = 2 / (N + 1)$

Verification: Mackey-Glass

The process to generate the data:

1. Randomly generate 3500 uniformly distributed data points between values 0 to 1. Let x be the series of these points. This series is further manipulated by the equation described below.
2. $x(t+1) = x(t) + \frac{0.2x(t-\tau)}{1+x(t-\tau)^{10}} - 0.1x(t)$
3. The value of τ in the above equation is set to 30. The above equation is used to change the random values, generated in step 1, adhering to the Mackey-Glass equation.
4. We then divide the series of 3500 points into 3 parts
5. 1st part is called the training set, the 2nd part is called validation set, 3rd set is called testing set with 72%, 8 % and 20% data in each set respectively.
6. With these sets being created, we then create the input matrix as discussed in our working paper.
7. We then use the test set as our ground truth values and predict the values for the same points using the model to see how well it performs.

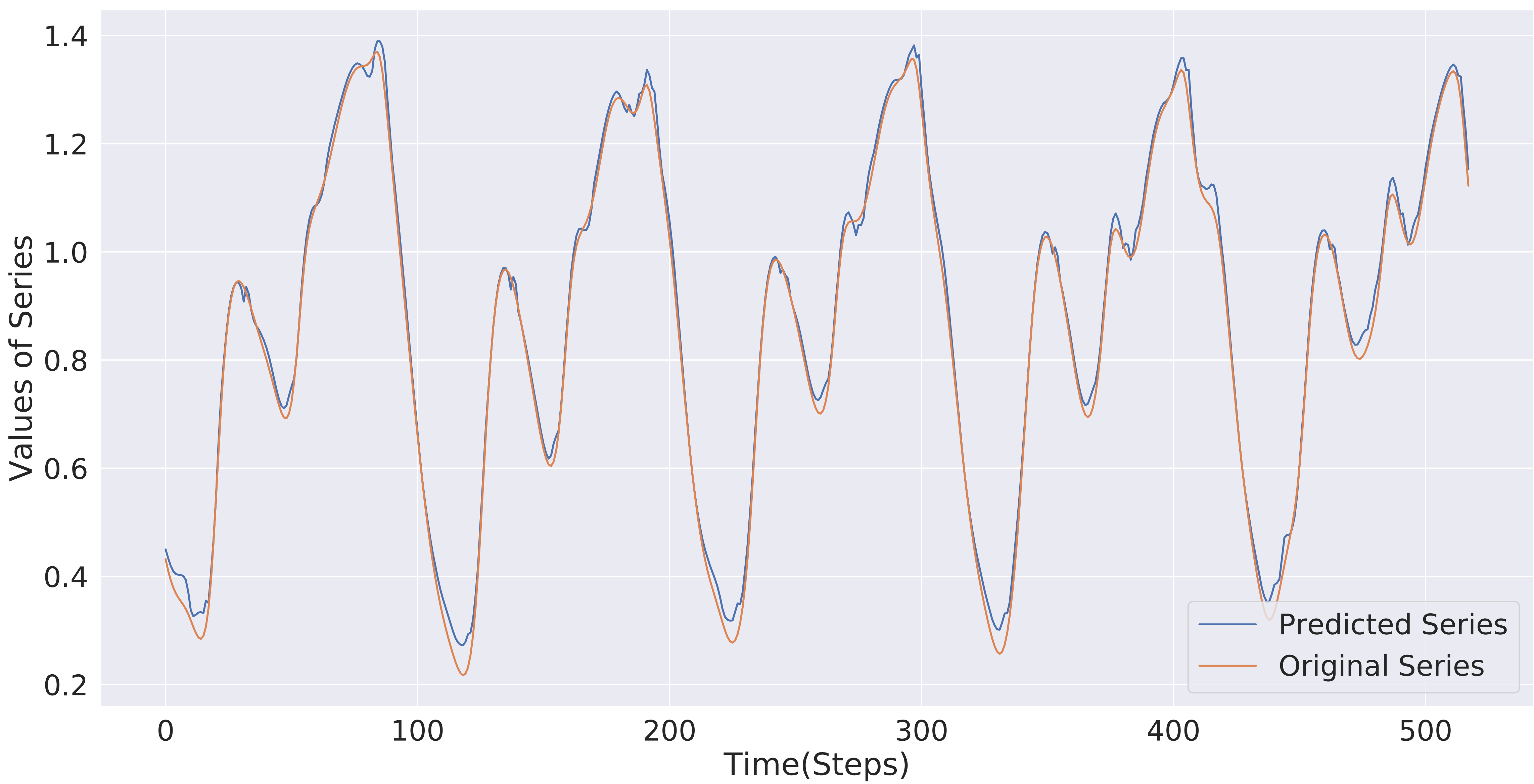


FIGURE 2: Mackey-Glass Time Series

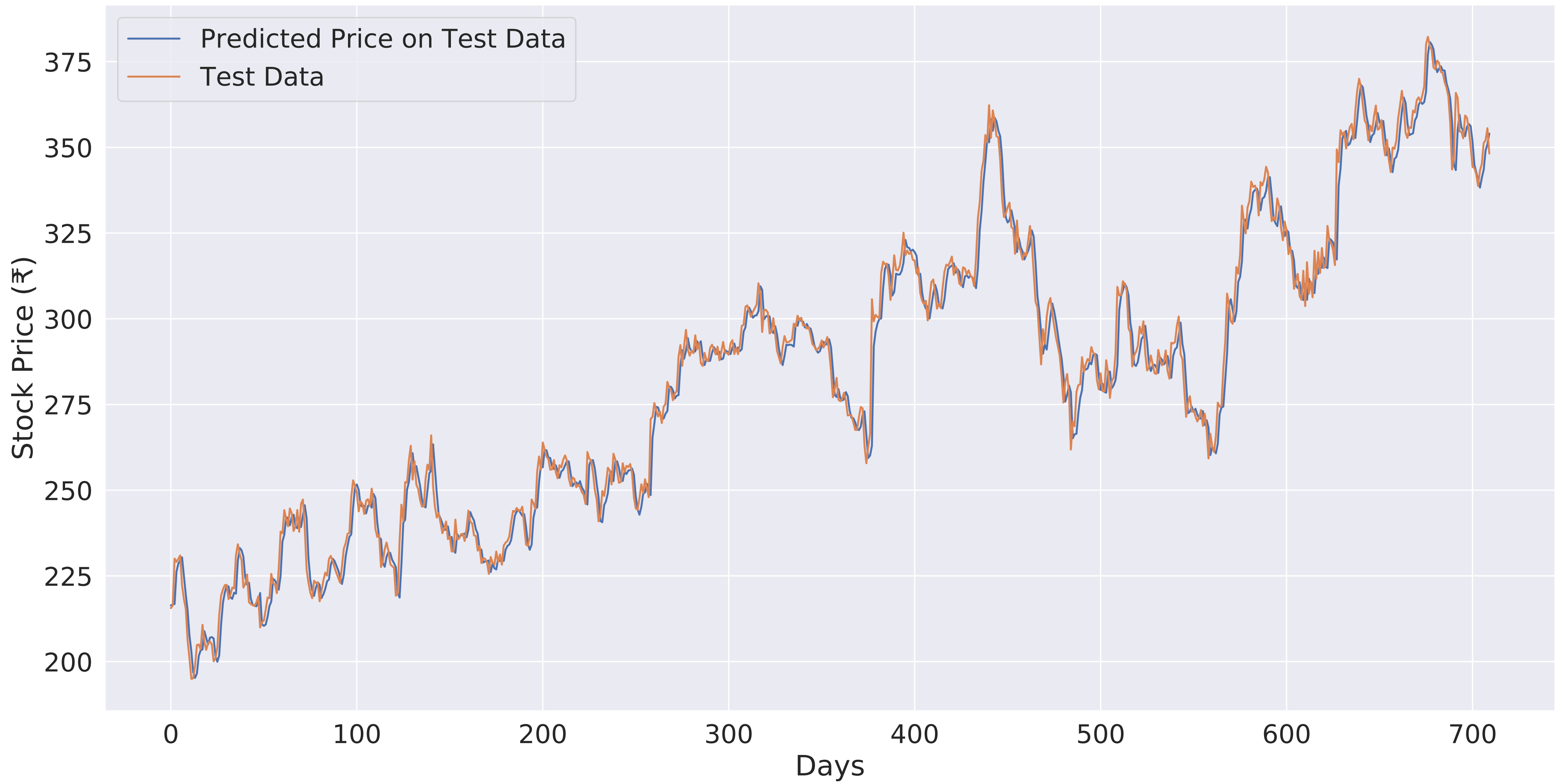
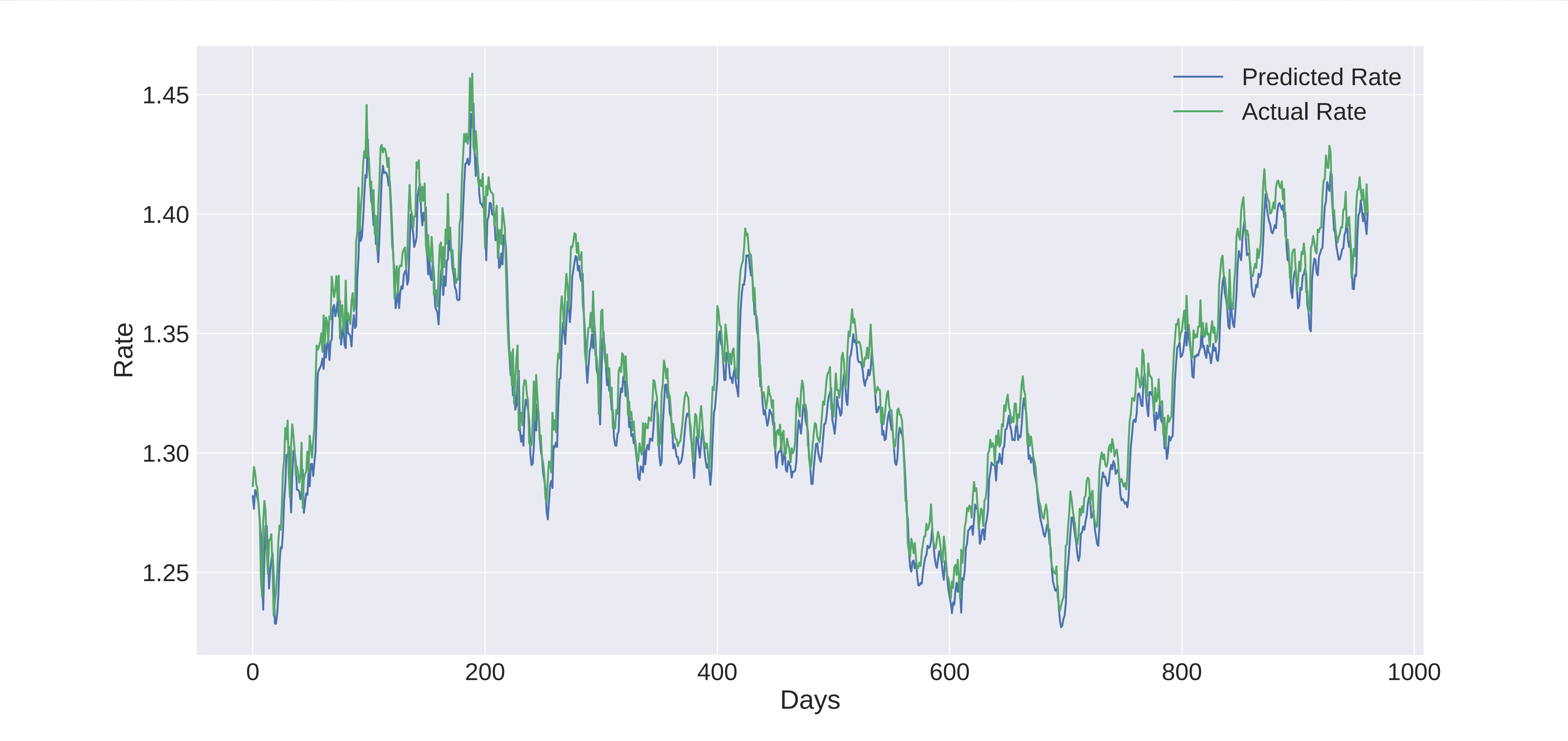


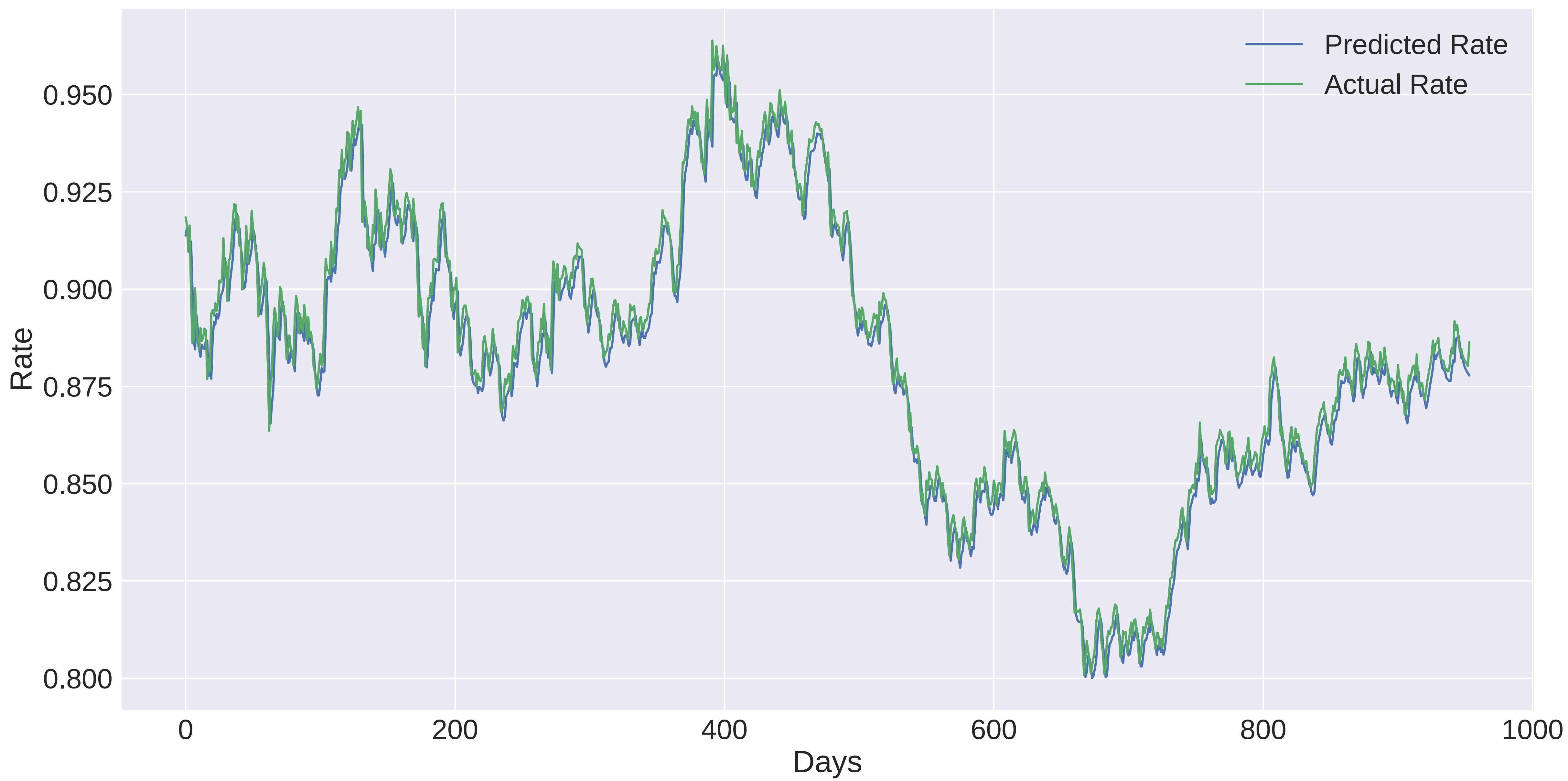
FIGURE 4: Predictions of ANN compared to original prices (test data) of ICICI Bank

FIGURE 5: Real time prediction of stock prices for ICICI Bank

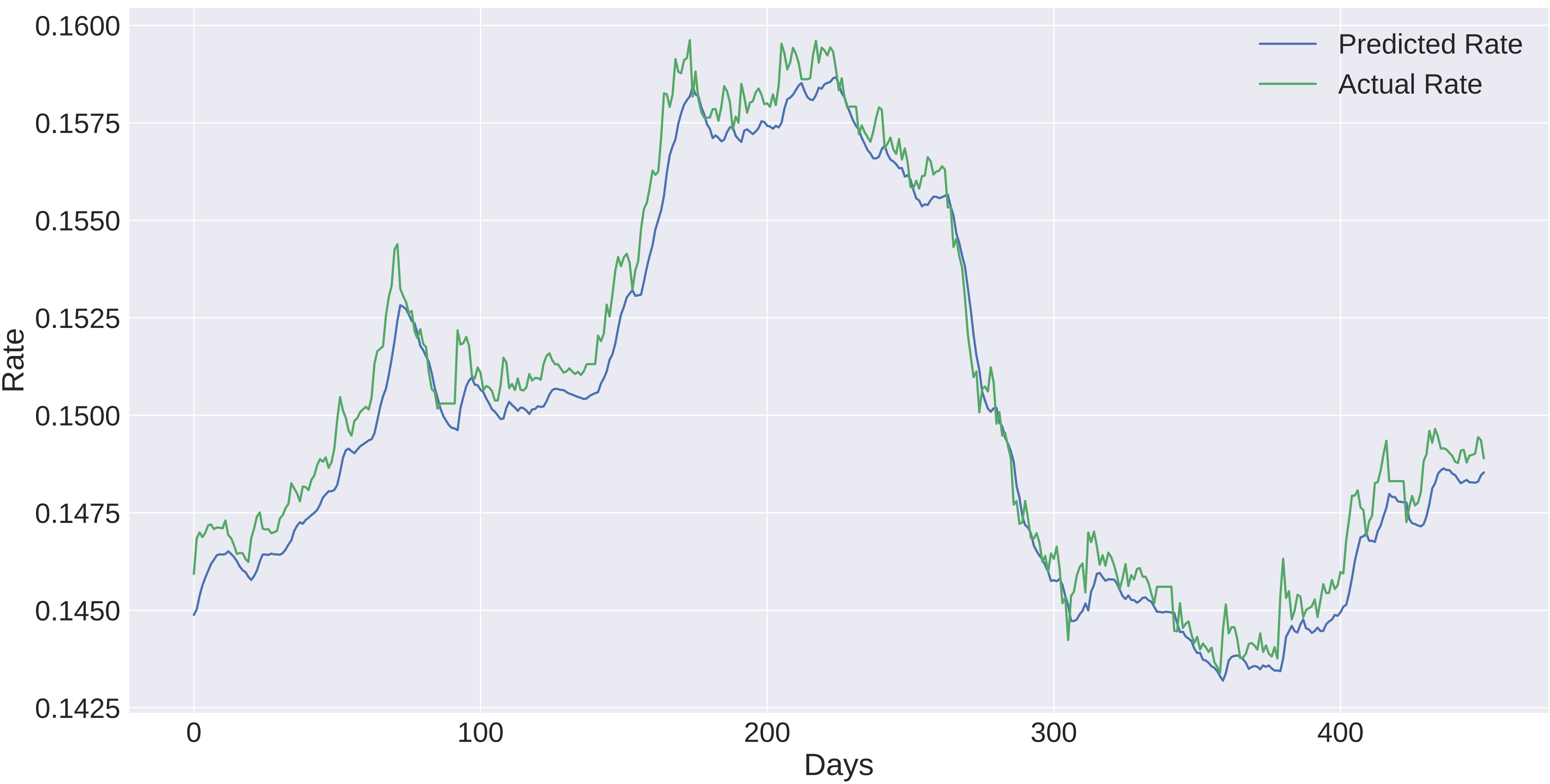
Currency Exchange



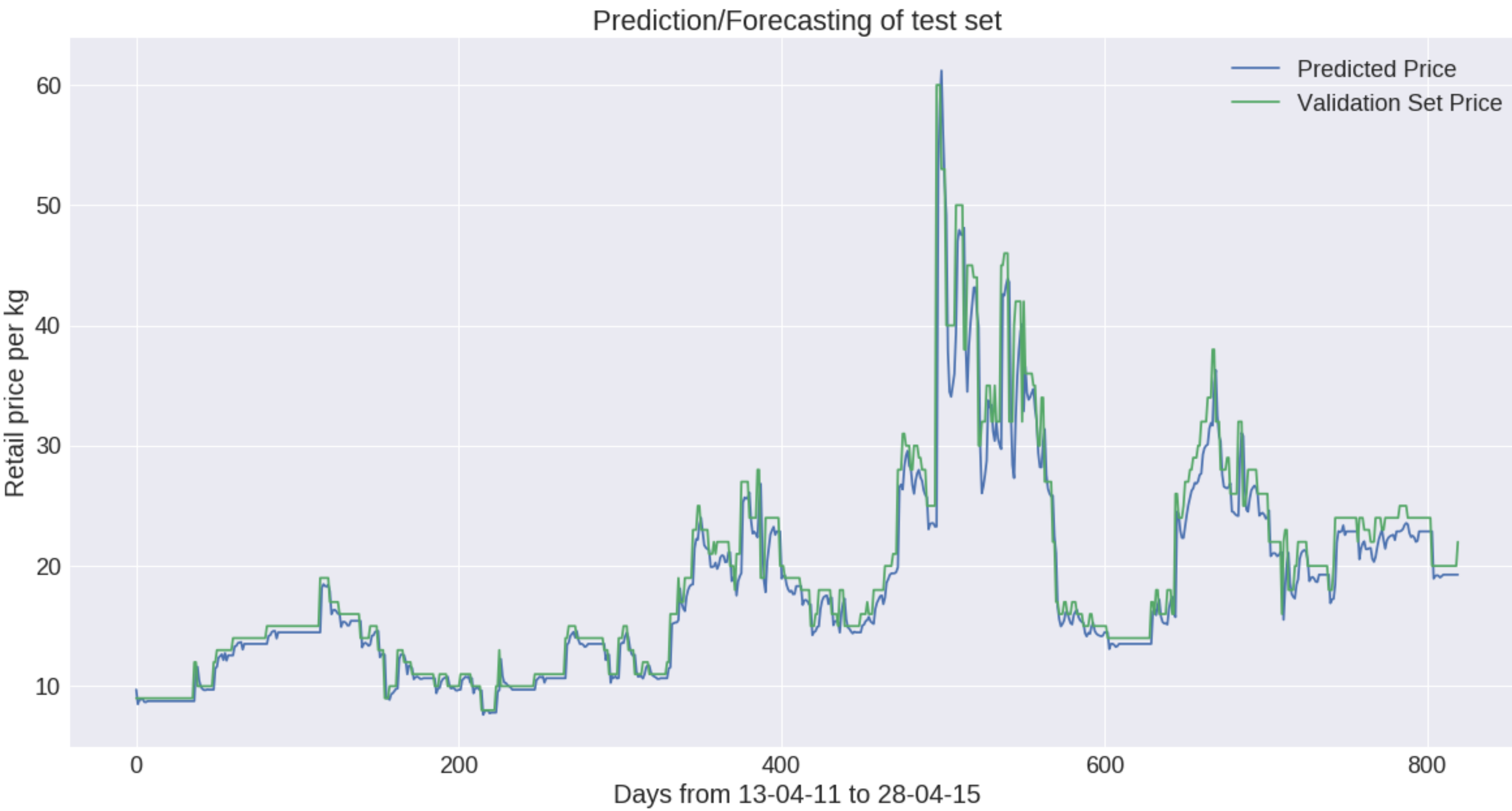
(a) Prediction of exchange rates of USD and AUD



(b) Prediction of exchange rates of USD and EUR



(c) Prediction of exchange rates of USD and CNY



Prediction/Forecasting of test set
Predicted Price
Validation Set Price
Retail price per kg
60
50
40
30
20
10
0
200
400
600
800
Days from 13-04-11 to 28-04-15

6 Conclusion

We approach the problem of forecasting time series using machine learning techniques. The proposed statistical features helped drew useful inferences on the time-series data. We took advantage of this and built a deep learning model on these statistical features. The model reconstructed unseen stock prices data using the learned weights. It also predicted the stock prices for the eighth day using the sliding window method. We extended this technique in forecasting the astronomical time series and exchange rates for USD - AUD, USD - EUR, and USD - CNY. The forecasting of light curve performed modestly but was unable to predict sudden fluctuations in the time series. The prediction of USD - CNY didn't work out well because functional data for training was not as much as the others. USD - AUD and USD - EUR prediction performed expertly.

In astronomy, a light curve is a graph of light intensity of a celestial object or region, as a function of time. The light is usually in a particular frequency interval or band

Financial Time Series Characteristics?

"Stylized facts"

Long time scale exists.
Long memory effects

Investors & Traders

Hurst exponent

Discovered by Hurst (1951), while investigating the discharge time series of the Nile River in the framework of the design of the Aswan High Dam, and found in many other hydrological and geophysical time series.

This behaviour is essentially the tendency of wet years to cluster into multi-year wet periods or of dry years to cluster into multi-year drought periods.

The terms 'Hurst phenomenon' and 'Joseph effect' (due to Mandelbrot, 1977, from the biblical story of the 'seven years of great abundance' and the 'seven years of famine') have been used as alternative names for the same behaviour.

The Encyclopedia of Water

Hydrological Persistence and the Hurst Phenomenon (SW-434)

Demetris Koutsoyiannis

Department of Water Resources, Faculty of Civil Engineering, National Technical University, Athens, Heroon Polytechneiou 5, GR-157 80 Zographou, Greece (*dk@itia.ntua.gr*)

The possible explanation of the long-term persistence must be different from that of the short-term persistence discussed above. This will be discussed later. However, its existence is easy to observe even in a time series plot, provided that the time series is long enough. For example, in Figure 1 (up) we have plotted one of the most well-studied time series, that of the annual minimum water level of the Nile river for the years 622 to 1284 A.D. (663 observations), measured at the Roda Nilometer near Cairo (Toussoun, 1925, p. 366-385; Beran, 1994). In addition to the plot of the annual data values versus time, the 5-year and 25-year averages are also plotted versus time. For comparison we have also plotted in the lower panel of Figure 1 a series of white noise (consecutive independent identically distributed random variates) with statistics same with those of the Nilometer data series. We can observe that the fluctuations of the aggregated processes, especially for the 25-year average, are much greater in the real world time series than in the white noise series. Thus, the existence of fluctuations in a time series at large scales distinguishes it from random noise.

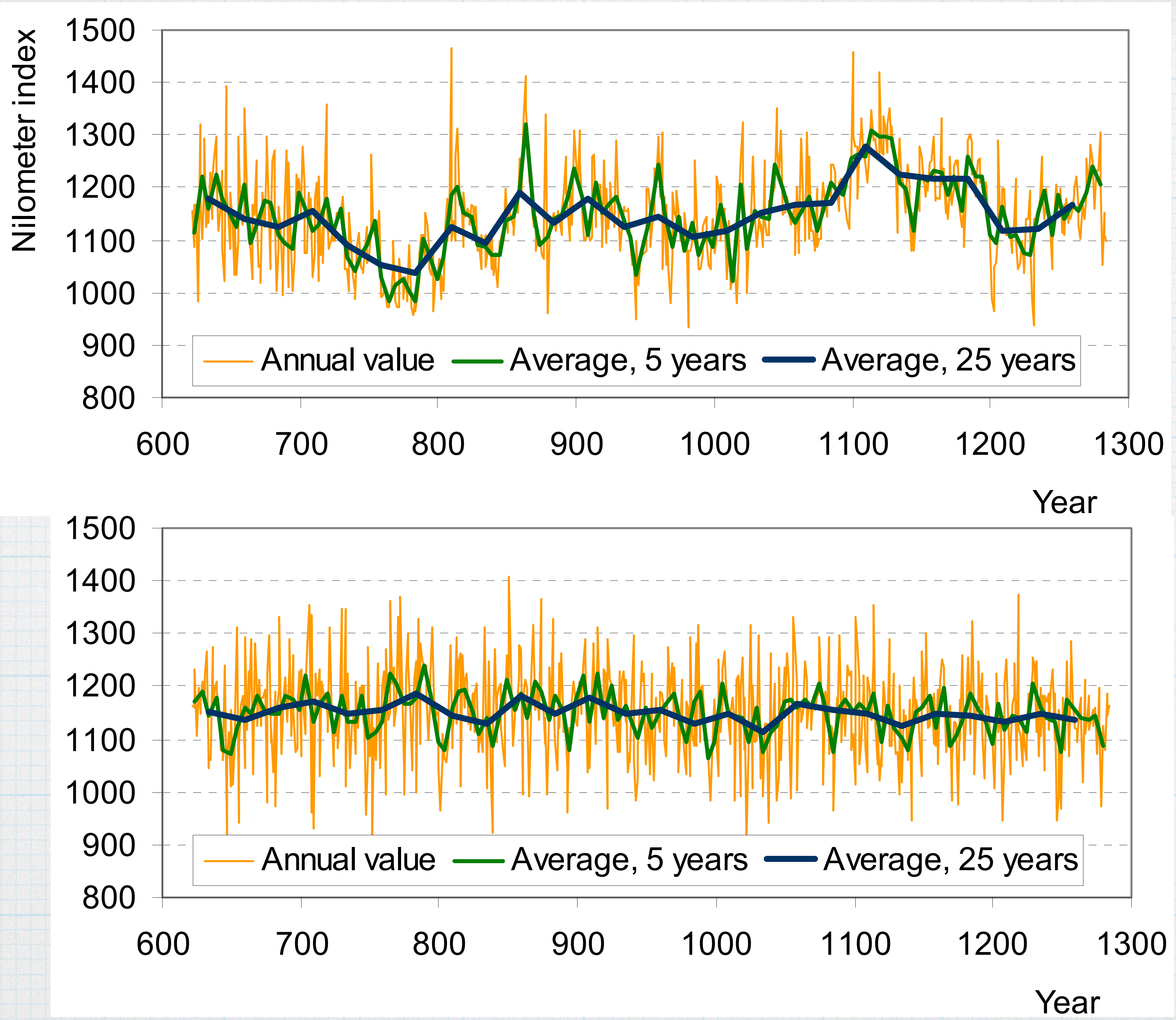


Figure 1 (Up) Plot of the Nilometer series indicating the annual minimum water level of the Nile River for the years 622 to 1284 A.D. (663 years); (down) a white noise series with same mean and standard deviation, for comparison.

The possible explanation of the long-term persistence must be different from that of the short-term persistence discussed above. This will be discussed later. However, its existence is easy to observe even in a time series plot, provided that the time series is long enough. For example, in Figure 1 (up) we have plotted one of the most well-studied time series, that of the annual minimum water level of the Nile river for the years 622 to 1284 A.D. (663 observations), measured at the Roda Nilometer near Cairo (Toussoun, 1925, p. 366-385; Beran, 1994). In addition to the plot of the annual data values versus time, the 5-year and 25-year averages are also plotted versus time. For comparison we have also plotted in the lower panel of Figure 1 a series of white noise (consecutive independent identically distributed random variates) with statistics same with those of the Nilometer data series. We can observe that the fluctuations of the aggregated processes, especially for the 25-year average, are much greater in the real world time series than in the white noise series. Thus, the existence of fluctuations in a time series at large scales distinguishes it from random noise.

What is, then, the difference between the "usual" stationary processes and long memory ones?

$$\downarrow$$

Annales de la Faculté des Sciences de Toulouse Vol. XV, n° 1, 2006
pp. 107–123

Long memory and self-similar processes[*]

Gennady Samorodnitsky [1]

The first thought that comes to mind is, obviously, about correlations. Suppose that $(X_n, n = 0, 1, 2, \ldots)$ is a stationary stochastic process with mean $\mu = EX_0$ and $0 < EX_0^2 < \infty$ (we will consider discrete time processes, but the corresponding formulations for stationary processes with finite variance in continuous time are obvious.) Let $\rho_n = \mathrm{Corr}(X_0, X_n)$, $n = 0, 1, \ldots$ be its correlation function. How does the correlation function of the "usual" stationary processes behave? It requires skill and knowledge to construct an example where the correlation function decays to zero (as lag increases) at a slower than exponentially fast rate. For example, the common linear (ARMA) processes, GARCH processes, finite state Markov chains all lead to exponentially fast decaying correlations. A process with correlations that are decaying slower than exponentially fast is, then, "unusual". If the correlations are not even absolutely summable, then the term "long memory" is often used. See Beran (1994).

Let Z_j, $j = 1, 2, \ldots$ be i.i.d. random variables with zero mean and non-zero finite variance. Choose a number $-1 < \rho < 1$, and an arbitrary (potentially random, but independent of the sequence Z_j, $j = 1, 2, \ldots$) initial state X_0. The AR(1) process X_n, $n = 0, 1, 2, \ldots$ is defined by

$$X_n = \rho X_{n-1} + Z_n, \quad n = 1, 2, \ldots. \tag{2.1}$$

It is elementary to check that the distribution of X_n converges to a limiting distribution, which is then automatically a stationary distribution of this simple Markov process. Choose the initial state X_0 according to this stationary distribution, which can be written in the form

$$X_0 = \sum_{j=0}^{\infty} \rho^j Z_{-j},$$

where $\ldots, Z_{-1}, Z_0$ are i.i.d. random variables independent of $Z_1, Z_2, \ldots$, and with the same distribution. Then the entire AR(1) process is already stationary, and

$$X_n = \sum_{j=0}^{\infty} \rho^j Z_{n-j}, \quad n = 0, 1, \ldots. \tag{2.2}$$

Notice that the correlation function of the stationary AR(1) process given by (2.2) is given by $\rho_n = \rho^n$, $n = 0, 1, 2, \ldots$, and it is exponentially fast decaying. In this sense the stationary AR(1) process is "usual". Notice also that the exponential rate of decay of correlations becomes slower as ρ approaches ± 1.

Of course, if ρ is exactly equal to -1 or 1, then the AR(1) process in (2.1) cannot be made stationary; in fact, if $\rho = 1$, then the AR(1) process is a random walk, which a is discrete-time equivalent of the Brownian motion. Therefore, in this case there is no real boundary layer between the "usual" stationary processes with exponentially fast decaying correlations and non-stationary ones. Nonetheless, even here, when ρ is close to -1 or 1, we may observe some features in the realizations of a stationary AR(1) process that remind us of non-stationary processes, such as a random walk. See the plots below.

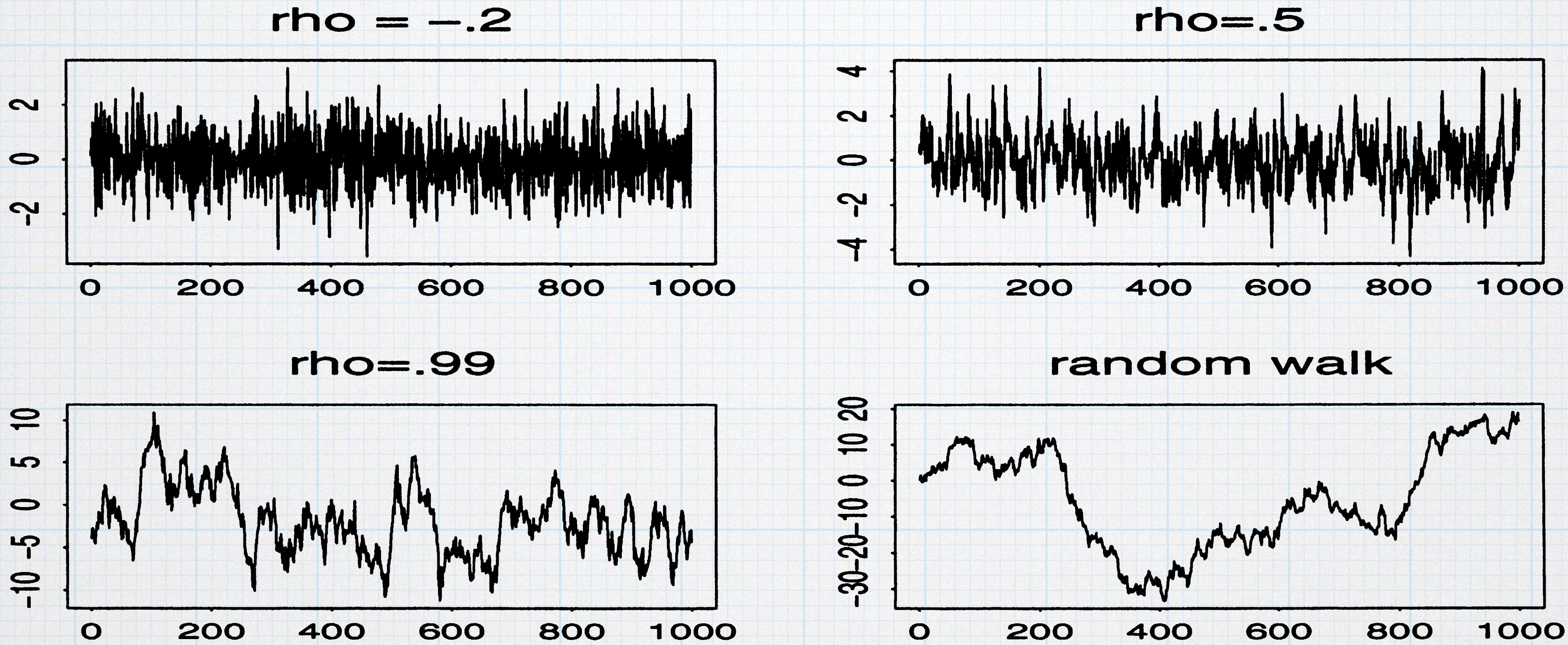


Figure 1. — Different AR(1) models and a random walk.

Overall, however, the class of AR(1) processes is too narrow to observe "unusual" stationary models that are close to being non-stationary.

3. Self-similar processes and long memory

Recall the definition of self-similarity: a stochastic process $(Y(t), t \geqslant 0)$ is called self-similar with exponent $H > 0$ of self-similarity if for all $c > 0$ the processes $(Y(ct), t \geqslant 0)$ and $(c^H Y(t), t \geqslant 0)$ have the same finite-dimensional distributions (*i.e.* scaling of time is equivalent to an appropriate scaling of space).

In applications a self-similar process is often a (continuous time) model for a cumulative input of a system in steady state, hence of a particular interest are self-similar processes reflecting this: processes with stationary increments. The common abbreviation for a self-similar process with stationary increments is SSSI (or H-SSSI if one wants to emphasize the exponent of self-similarity). We refer the reader to Samorodnitsky and Taqqu (1994) and Embrechts and Maejima (2002) for information on the properties of self-similar processes, including some of the facts presented below.

Suppose that $(Y(t), t \geqslant 0)$ is an H-SSSI process with a finite variance. Since we trivially have $Y(0) = 0$ a.s., we see that

$$E(Y(t) - Y(s))^2 = E\Big(Y(t-s) - Y(0)\Big)^2 = EY^2(t-s) = (t-s)^{2H} EY^2(1)^2$$

for all $t > s \geqslant 0$, and so

$$\operatorname{Cov}(Y(s), Y(t)) = \frac{1}{2}\Big[EY^2(t) + EY^2(s) - E(Y(t) - Y(s))^2\Big] \tag{3.1}$$

$$= \frac{EY^2(1)}{2}\Big[t^{2H} + s^{2H} - (t-s)^{2H}\Big].$$

Assuming non-degeneracy ($EY(1)^2 \neq 0$), the expression in the right hand side of (3.1) turns out to be nonnegative definite if and only if $0 < H \leqslant 1$, in which case it is a legitimate covariance function.

In particular, for every $0 < H \leqslant 1$ there is a unique zero mean Gaussian process whose covariance function is consistent with self-similarity with exponent H and stationary increments and, hence, is given by (3.1). Conversely, a Gaussian process with a covariance function given by (3.1) is, clearly, both self-similar with exponent H and has stationary increments. That is, for every $0 < H \leqslant 1$ there is a unique (up to a global multiplicative constant) H-SSSI zero mean Gaussian process. This process is called Fractional Brownian Motion (FBM), and will be denoted by $(B_H(t), t \geqslant 0)$. It is trivial to check that for $H = 1$, $E\big(tB_H(1) - B_H(t)\big)^2 = 0$ for all $t \geqslant 0$, which means that the process is a straight line through the origin and random normal slope. The interesting and nontrivial models are obtained when $0 < H < 1$, which is what we will assume in the sequel.

The increment process $X_n = B_H(n+1) - B_H(n)$, $n \geqslant 0$ of an FBM is a stationary process called Fractional Gaussian noise (FGN). An immediate conclusion from (3.1) is that

$$\rho_n = \mathrm{Corr}(X_0, X_n) \sim 2H(2H-1)n^{-2(1-H)}$$

as $n \to \infty$. Therefore, the correlation function of an FGN with $1/2 < H < 1$ satisfies (2.4) with $d = 2(1-H) < 1$. Fractional Gaussian noises with $H > 1/2$ are commonly viewed as long range dependent. Since the process is a Gaussian one, this is not particularly controversial, because the covariance on whose behavior the term "long range dependent" hinges here, determines the structure of the process.

$$\rho_n = n^{-d}L(n),\ n = 0, 1, 2, \ldots, \tag{2.4}$$

L is slowly varying at infinity

Fractional Brownian Motion with $H > 1/2$ was used by Mandelbrot and Van Ness (1968) and Mandelbrot and Wallis (1968) to give a probabilistic model consistent with an unusual behaviour of water levels in the Nile river observed by Hurst (1951), and it was noted already there that the realizations of long range dependent Fractional Gaussian noises may exhibit apparently obvious non-stationarity. For example, as one compares the four plots on Figure 2, it appears that, for larger values of H, the plots tend to indicate changing "level", or the mean values, of the process at different time intervals. The phenomenon is dramatic in the case $H = .9$. This is, of course, in spite of the fact that the process is stationary and the mean is always zero.

Notice that for $0 < H < 1/2$ the correlations of a FGN tend to be negative, and for $H = 1/2$ the FGN is simply an i.i.d. sequence.

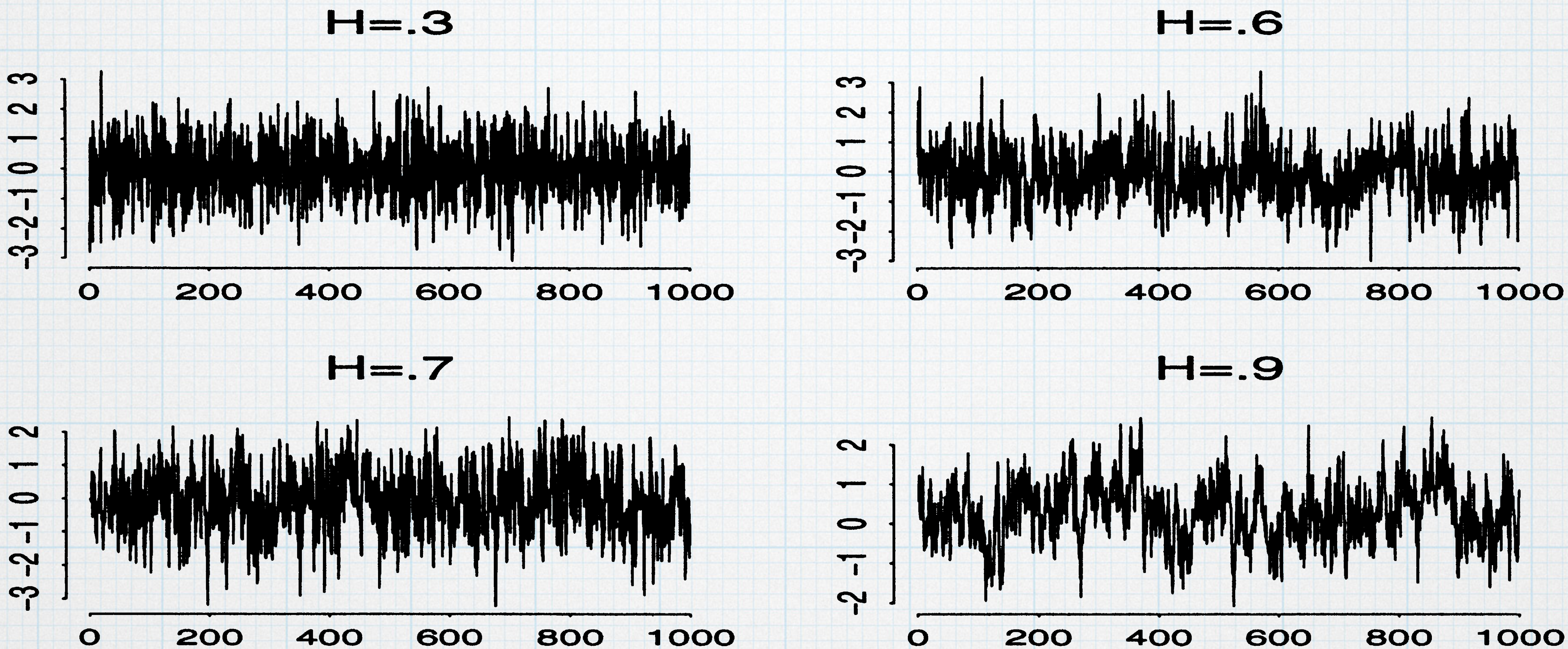


Figure 2. — Different Fractional Gaussian noises.

Moral

H < 0.5 Short range

=0.5 iid ( )

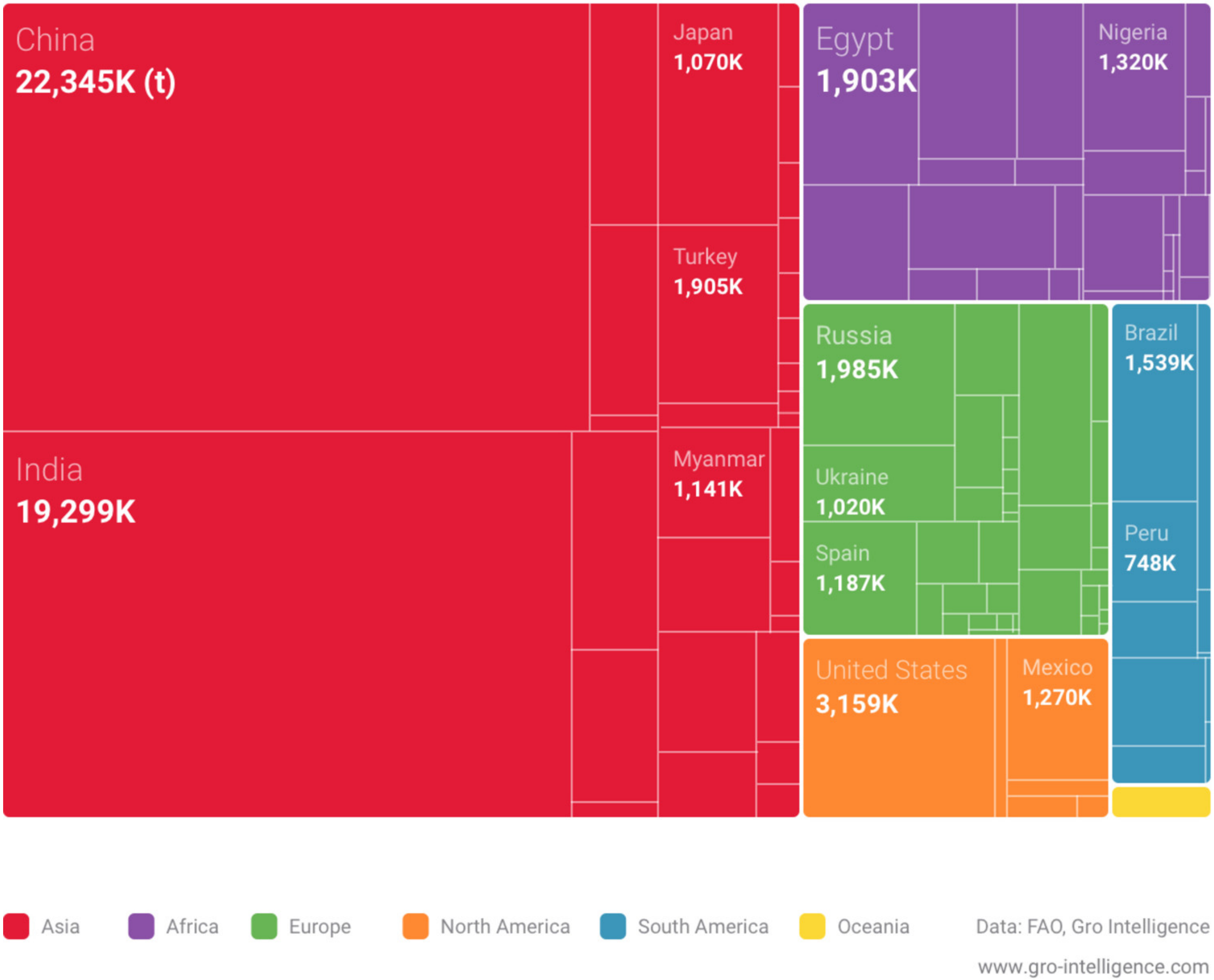
> 0.5 Long range

Onion Prices

Onions

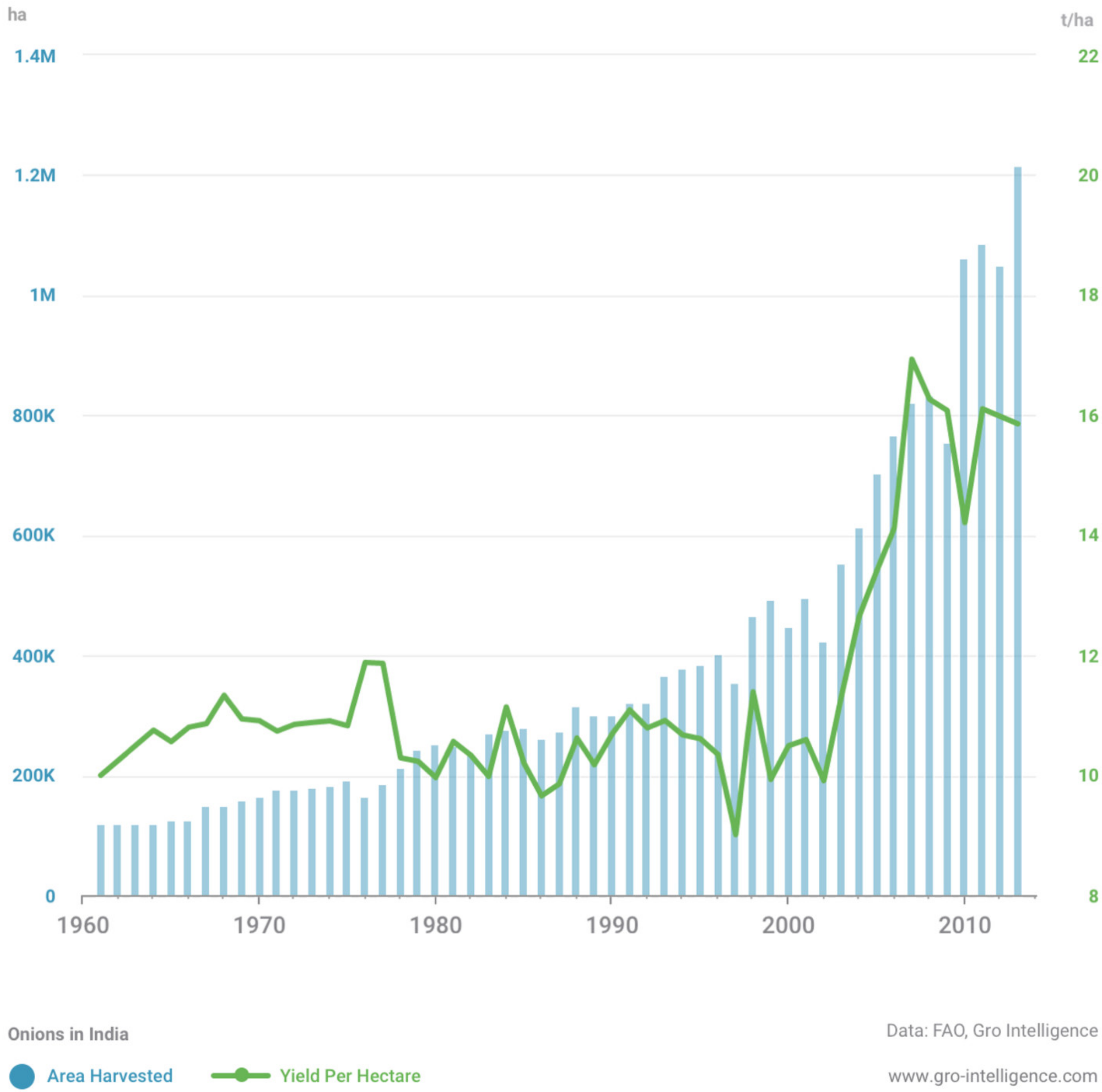
India is the world’s second-largest producer of onions, after China. Even though India consumes and produces a great deal of onions, it has had considerable trouble smoothing out eye-watering price volatility. In various major Indian cities, retail prices of onions more than quadrupled between May 2013 and October 2013, before falling back at the beginning of the following year; a similar spike happened between May 2015 and September 2015, in which retail prices nearly doubled before falling back down.

Knowing our onions



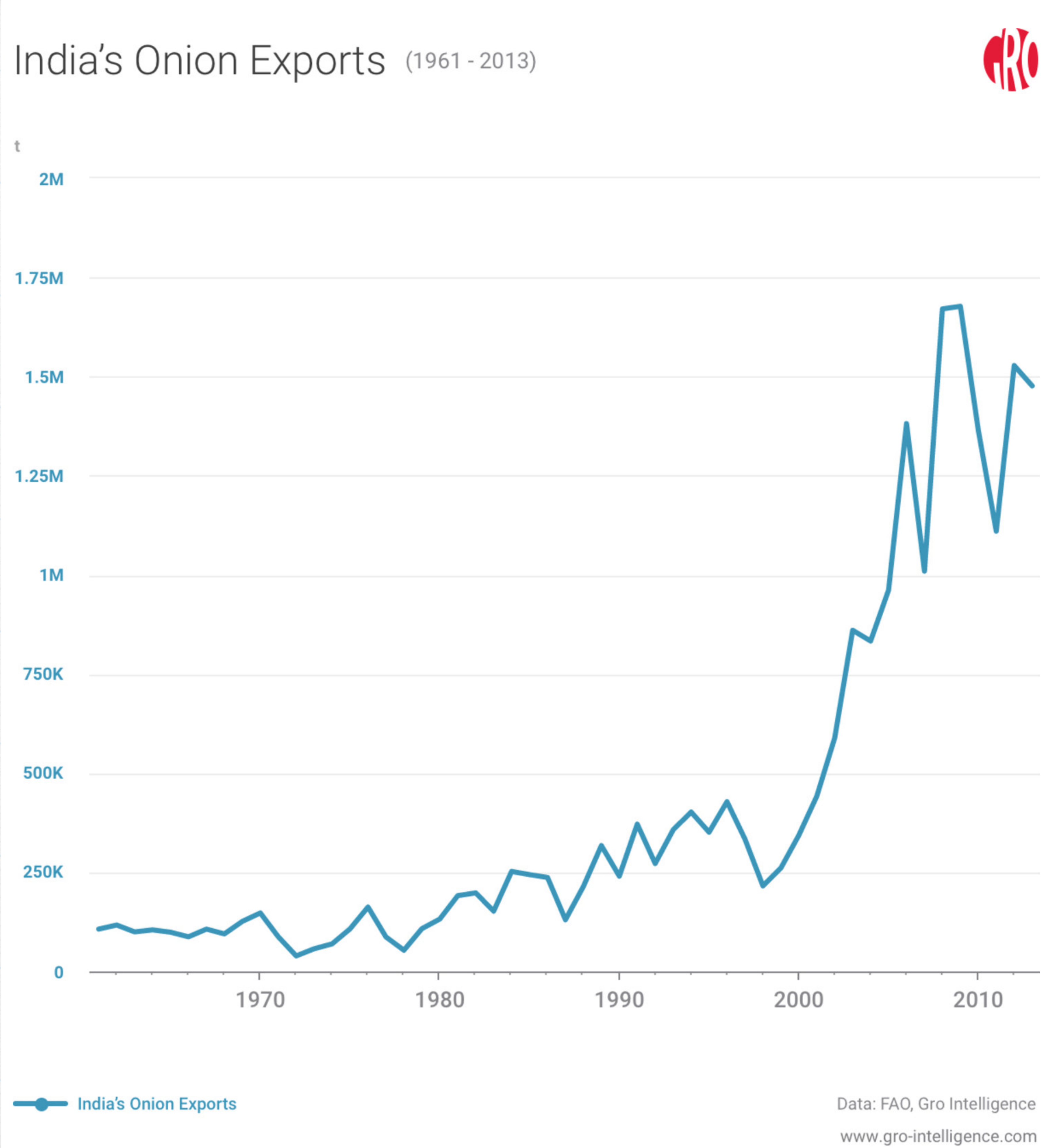
(https://clews.gro-intelligence.com/#/welcome)

India's Onion Yield and Area Harvested (1960 - 2013)

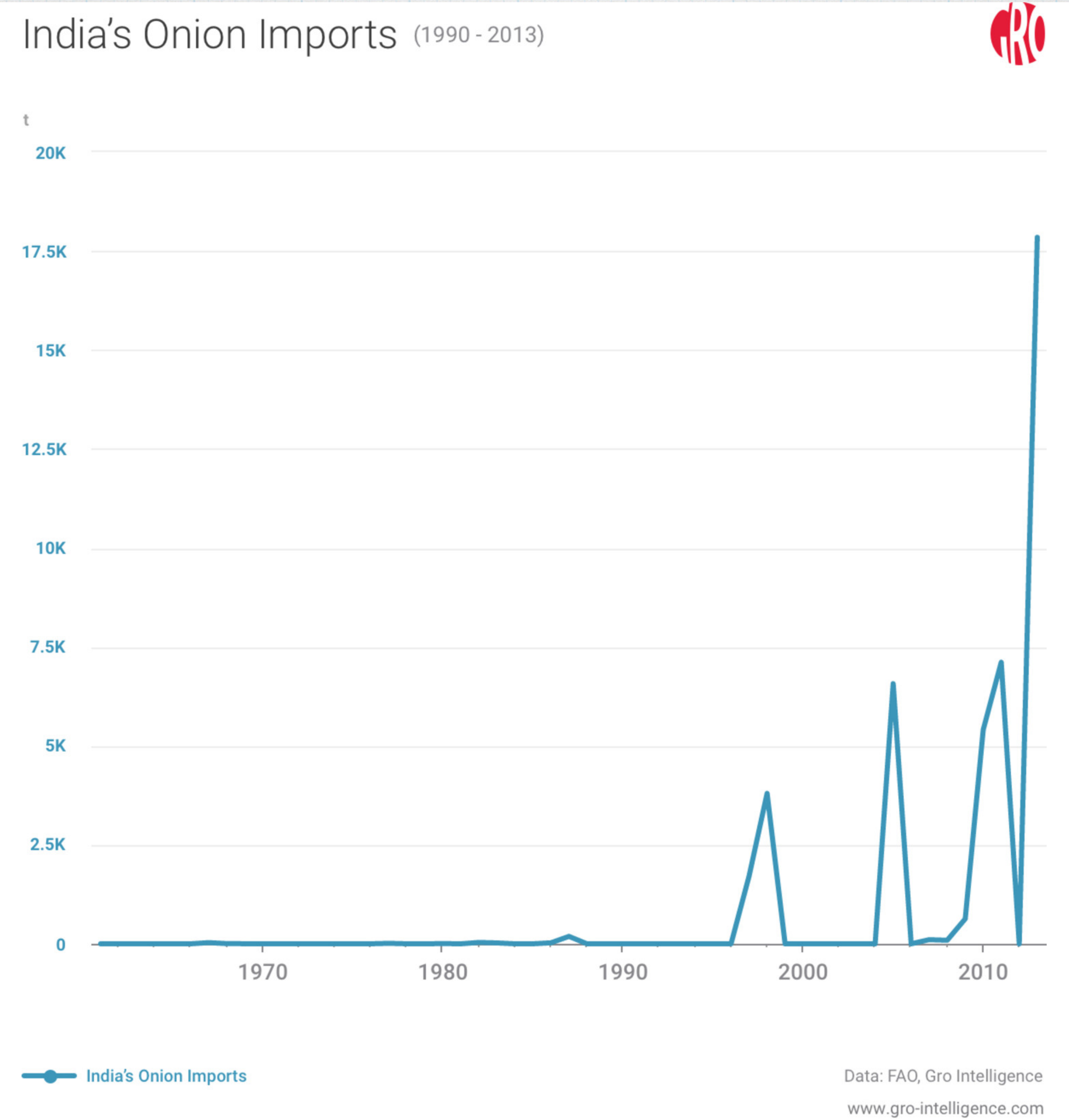


Data: FAO, Gro Intelligence

www.gro-intelligence.com



India's Onion Exports (1961 - 2013)
t
2M
1.75M
1.5M
1.25M
1M
750K
500K
250K
0
1970
1980
1990
2000
2010
India's Onion Exports
Data: FAO, Gro Intelligence
www.gro-intelligence.com


India's Onion Imports (1990 - 2013)
t
20K
17.5K
15K
12.5K
10K
7.5K
5K
2.5K
0
1970
1980
1990
2000
2010
India's Onion Imports
Data: FAO, Gro Intelligence
www.gro-intelligence.com

In response to various price hikes and supply shortages, and given that the government severely restricts imports, the government has favored increasing domestic supply through export controls. These policies usually have had mixed effectiveness, and the government may be able to reduce volatility by making greater investments in increasing production and improving storage over the long term.

Although India's government highly values a stable supply of onions, it permits very few imports to enter the country. Trade in onions, especially imports, is highly regulated. Even when the government foresees that weather will reduce the size of a harvest, it tends to permit imports only after prices start to rise. The usual reason the government offers for restricting imports is that it wants to narrow its current-account deficit.

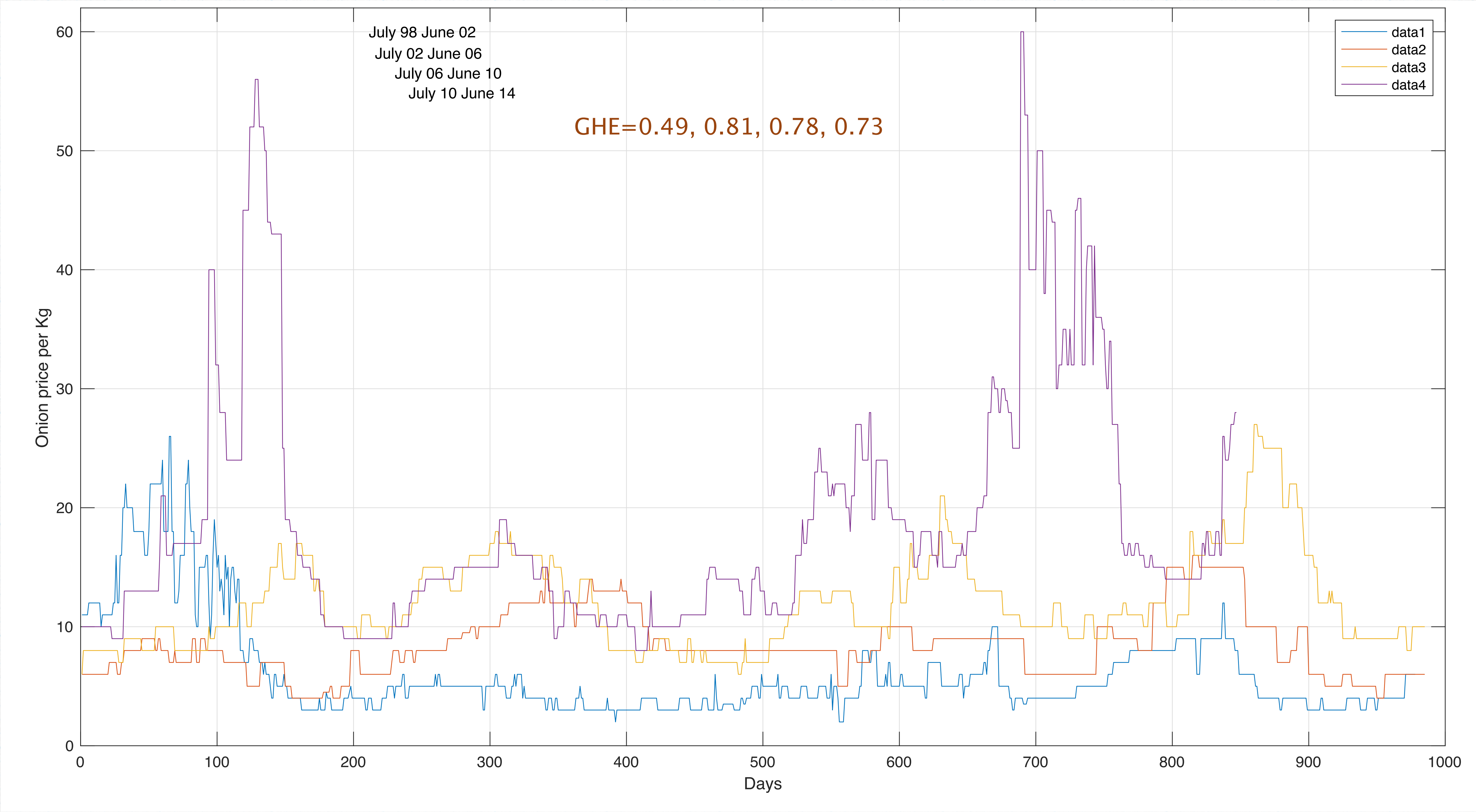
Imports of onions require government approval and are organized by state-run trading companies, also known as canalizing agencies. When prices are high, these agencies are authorized to seek overseas suppliers to increase domestic supply. Even if these companies manage to find onion supplies on short notice, shipments have historically arrived too late to prevent a price spike.

The Indian government more frequently resorts to restricting exports in order boost domestic supply. In addition to short-term bans, the government more often relies on minimum export prices (MEP) to reduce the amount that gets shipped overseas. It has raised MEP a number of times over the last decade, including in 2012, 2013, 2014, and 2015. Occasionally, these MEPs serve the same function as export bans, given that producers aren't able to find overseas buyers at prices well above market prices.

These interventions on distribution have had mixed effects as they have been usually enacted too late to control for price hikes. Even when the government decides to import onions, the supplies rarely arrive in time to relieve consumers. Instead, these interventions may hurt long-term onion production, blocking producers from taking advantage of profitable exports and higher margins. Onions from India are generally well-regarded in overseas markets, and command (http://sfacindia.com/PDFs/Onion%20&%20Potato%20Baseline%20Report.pdf) a price premium over onions from Pakistan and China. When the government acts constantly to keep prices of onions down, producers have less incentive to make significant investments in increasing production.

Instead of resorting to trade restrictions, the government might do better for consumers by investing in storage and distribution infrastructure. Increasing storage capacity could help reduce some of the volatility in the markets and create a buffer stock of the vegetable that can be built up during peak supply and steady prices until the next harvest reaches the market. Further, modern storage infrastructure that prevents condensation which leads to rot would reduce post-harvest losses, which at 40 percent, are significant. As part of infrastructure development, investment in greater irrigation capacity would reduce farmers dependence on seasonal rains. In addition, the establishment of more processing facilities would allow more onions to be processed during gluts to later be consumed during shortages. Turning the less aesthetically pleasing onions into pastes and sauces would reduce the waste of otherwise fine onions.

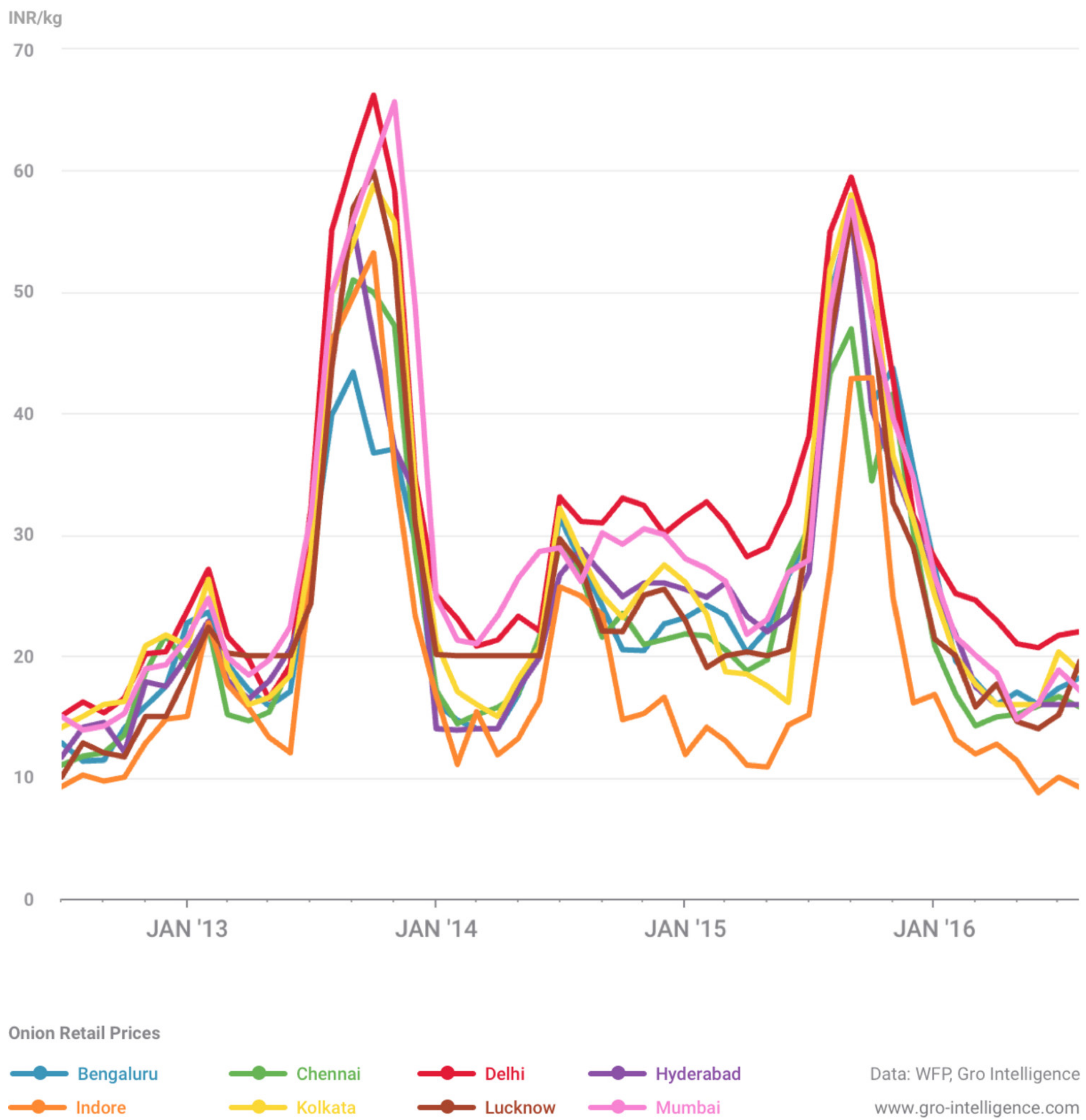
//data.gov.in/



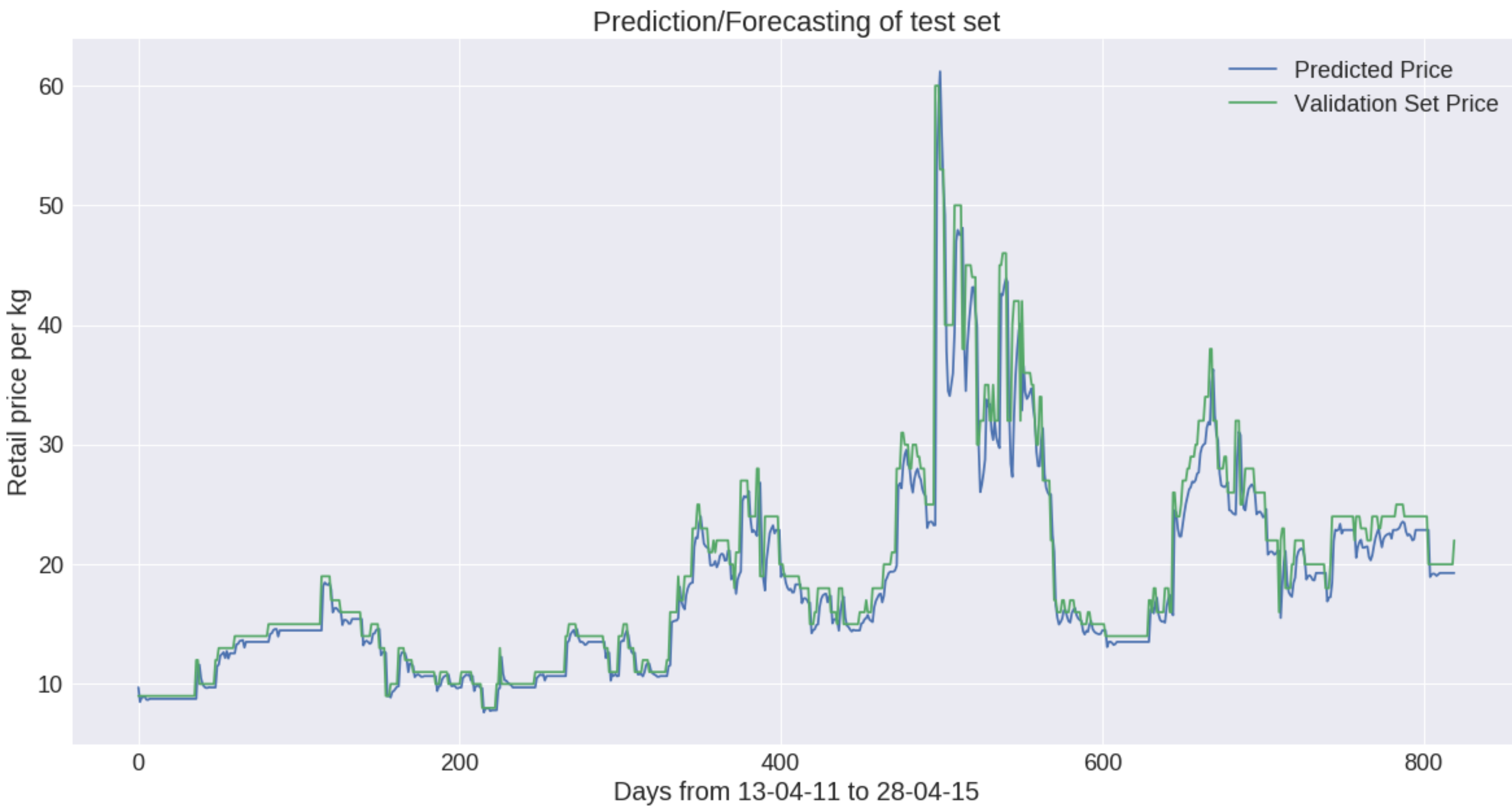
July 98 June 02
July 02 June 06
July 06 June 10
July 10 June 14
GHE=0.49, 0.81, 0.78, 0.73
data1
data2
data3
data4
Onion price per Kg
Days
0
10
20
30
40
50
60
0
100
200
300
400
500
600
700
800
900
1000

Retail Price of Onions in Major Indian Cities

(July 2012 - August 2016)



(https://clews.gro-intelligence.com/#/welcome)



Prediction/Forecasting of test set
Predicted Price
Validation Set Price
Retail price per kg
10
20
30
40
50
60
0
200
400
600
800
Days from 13-04-11 to 28-04-15

Incorporate these features into the proposed algorithm

Thank You