GROUNDWATER QUALITY IN INDIA

An Analysis of the Primary Factors contributing to Ground Water Pollution in India

GROUP DETAILS
Shoumik Goswami
Vanessa D' Souza
Vyshak Reghukumar

Contents

Introduction

Groundwater pollution is a major environmental issue in India. The largest source of water pollution in India is untreated sewage. Other sources of pollution include agricultural runoff and unregulated small scale industry. Most rivers, lakes and surface water in India are polluted.

For this research, we have analyzed the water quality data of groundwater sourced from various water stations across the country. Groundwater pollution (also called groundwater contamination) occurs when pollutants are released to the ground and make their way down into groundwater. It can also occur naturally due to the presence of a minor and unwanted constituent, contaminant or impurity in the groundwater, in which case it is more likely referred to as contamination rather than pollution.

Problem Statement

1. As a whole, which areas in India has a lot of water degradation issues over the years.
2. Explore if there are any repetitive patterns of water quality degradation in the same area for multiple years.
3. Understand the factors that are leading to the lower water quality.
4. Predict if there are going to be water quality issues in areas that are not affected right now based on the water quality data that is available right now.

Data

The dataset to be used for the study has been downloaded from: https://data.gov.in/catalog/water-quality-india-2014

The dataset was released by National Data Sharing and Accessibility Policy (NDSAP). The combined dataset contains the following:

1. Number of locations: 447
2. Number of Variables to determine water quality: 6
3. Number of states covered in the dataset: 22

Sl. No.	Variable	Definition
1	Temperature	Temperature of the groundwater
2	pH	pH level of the groundwater (acceptable level: 6.5-8.5)
3	Conductivity	Ability of groundwater to conduct current
4	B.O.D. (Biochemical Oxygen Demand)	Amount of dissolved oxygen in the groundwater
5	Nitrate (N+) Nitrite (N)	N+ and N levels
6	Fecal Coliform	Level of fecal contamination

Methodology

Water Quality Index (WQI) is a very useful and efficient method which can provide a simple indicator of water quality and it is based on some very important parameters. In current study, Water Quality Index (WQI) was calculated by using the Weighted Arithmetic Index method. In this model, different water quality components are multiplied by a weighting factor and are then aggregated using simple arithmetic mean.

For assessing the quality of water in this study, firstly, the Quality Rating Scale (Qi) for each parameter was calculated by using the following equation;

$$Q_i = \frac{100 * (V_{actual} - V_{ideal})}{(V_{standard} - V_{ideal})}$$

Where,

Q_i = Quality rating of i-th parameter for a total of n water quality parameters

V_{actual} = Actual value of the water quality parameter obtained from laboratory analysis

V_{ideal} = Ideal value of that water quality parameter can be obtained from the standard Tables.

V_{ideal} for pH = 7 and for other parameters it is equaling to zero, but for DO V_{ideal} = 14.6 mg/L

$V_{standard}$ = Recommended WHO standard of the water quality parameter.

Next, after calculating the quality rating scale (Qi), the Relative (unit) weight (Wi) was calculated by a value inversely proportional to the recommended standard (Si) for the corresponding parameter using the following expression;

$$W_i = \frac{1}{S_i}$$

Where,

W_i = Relative (unit) weight for nth parameter

S_i = Standard permissible value for nth parameter

I = Proportionality constant.

That means, the Relative (unit) weight (WI) to various water Quality parameters are inversely proportional to the recommended standards for the corresponding parameters.

Finally, the overall WQI was calculated by aggregating the quality rating with the unit weight linearly by using the following equation:

$$WQ_i = \frac{\sum Q_i W_i}{\sum W_i}$$

Where,

Q_i = Quality rating

W_i = Relative weight in general

WQI is defined for a specific and intended use of water. In this study the WQI was considered for human consumption or uses and the maximum permissible WQI for the drinking water was taken as 100 score.

Table 1: Calculation of Si and Wi for the various parameters

Parameters	**Si**	**Wi**
pH	8.5	0.11
Conductivity	250	0.004
Temperature	25	0.1
BOD	6	0.11
Nitrate	200	0.1
Fecal Coliform	20	0.16

Table 2: Water Quality Classification Based On WQI Value

S. No	**WQI value**	**Water quality**
1	<50	Excellent
2	50-100	Good
3	100-250	Poor
4	250-350	Very poor
5	>350	Water unsuitable for drinking

Besides WQI, Time Series Analysis was also done, which includes ARIMA and Holt-Winter's exponential smoothing, to predict the pollution trends of the various states in future.

Time series data often arise when monitoring industrial processes or tracking corporate business metrics. Time series data may be defined as an ordered sequence of values of a variable at equally spaced time intervals.

ARIMA models are a popular and flexible class of forecasting model that utilize historical information to make predictions. This type of model is a basic forecasting technique that can be used as a foundation for more complex models.

For a time series that can be described using an additive model with increasing or decreasing trend and seasonality, Holt-Winters exponential smoothing can be used to make short-term forecasts. Holt-Winters exponential smoothing estimates the level, slope and seasonal component at the current time point.

The data was segregated based on the states and the WQI values of every location was calculated. The respective time series models were created for those states which had data for a continuous period of 3 years or more. After creating the time series models of every parameter, plots were generated to understand the trends and patterns in the data.

In order to use ARIMA models for prediction, it was necessary to have a stationary data and at-least 38 data-points. A stationary series has no trend, its variations around its mean have a constant amplitude, and it wiggles in a consistent fashion, i.e., its short-term random time patterns always look the same in a statistical sense. Since the number of data-points were considerably less, ARIMA models could not be used for the analysis.

Triple Exponential Smoothing, also known as the Holt-Winters method, is one of the many methods or algorithms that can be used to forecast data points in a series, provided that the series is “seasonal”, i.e. repetitive over some period. Holt-Winter is used for exponential smoothing to make short-term forecasts by using “additive” or “multiplicative” models with increasing or decreasing trend and seasonality. Smoothing is measured by beta and gamma parameters in Holt’s model. In cases where no trend was observed in the data, the value of beta was taken as false.

The entire procedure needs to be repeated for every state in contained in the dataset, provided there are not missing years in the series. An automatic script in R was generated to cater to this need. The script required a generalised code which can select the required dataset of each state and forecast the values of the parameters for the next consecutive year. Once all the possible values were forecasted for every parameter of each state respectively, the WQI values were calculated for the future time periods.

Tableau was used on the new dataset generated and geo-spatial plots were created that resulted in a timeline of change of ground water quality state-wise from the period 2006 to 2017. The geospatial plots highlight how the ground water quality varies over time for each state. Pivots were created in MS-Excel using the dataset generated to understand how the ground water quality was being affected by the change of parameters over time.

Results

Water quality data for yearly data across various states in India was successfully collected, transformed, modeled, predicted, visualised and analysed with a heady mix of Excel, R, Rmd, Tableau and statistical modeling techniques.The water quality data from Government sites was observed over the time range of 2006 to 2017. The 6 parameters, pH, Conductivity, Temperature, BOD, Nitrate, Fecal Coliform were analysed and time series was used to predict future yearly values and subsequently calculate WQI for the next 3 periods. To predict values, Holt's winter exponential smoothing was selected based on the data set available. To understand data for various states, generated automated state wise reports on a single click with the help of R and RMarkdown. Time series of the yearly data of parameters for every location were analyzed and the WQI values were calculated. To visualise data at country level, geospatial plots were created in Tableau and compared with the permissible values as stated by WHO. These Reports, geospatial plots and graphs were analysed and can be used to understand and improve water quality in India.

Discussion of Results

Andhra currently has excellent WQI (2014 – 32); but it is steadily increasing (2015 – 57, 2016 -75, 2017 – 93), implying that groundwater in Andhra is getting polluted.

1. pH is has been on the higher side and shows increasing trend, implying alkalinity.
2. Conductivity falling down, implying that salinity is decreasing.
3. BOD is also showing steady growth, implying that it contains potentially harmful bacteria or bacteria producing undesirable physical characteristics such as taste or odor.
4. Nitrate and Nitrite levels also found to be high, though they are generally reducing.
5. Fecal coliform is also showing growing trends, indicating contamination by human sewage or animal droppings.

Assam had alarmingly high WQI in 2006 and 2007, though they've drastically improved in 2008 (possibly due to availability of more data). However, there is still risk of increase in WQI in upcoming years (2015 – 197, 2016 – 296, 2017 – 387).

1. pH level is in acceptable levels, though recent years indicate possibility of acidity.
2. Conductivity has been increasing in recent years, implying increasing salinity.
3. BOD shows irregular trends, though they have been on the lower side in recent years.
4. Nitrate and Nitrite levels increasing explosively, which is a matter of concern.
5. Fecal coliform has shown tremendous improvement, though it is steadily rising which is a matter of concern.

Goa has excellent WQI (2014 – 50); however it shows increasing trends, implying possible contamination.

1. pH is generally on the lower side, indicating high acidity. However it has improved in recent years.
2. Conductivity has increased manifold, indicating immense salinity.
3. BOD is showing a slight decline.
4. Nitrate and nitrite levels are generally good, though they showed a slight increase. However the latest data (2014) indicates a lower level.
5. Fecal coliform has increased greatly in recent years, though the latest data (2014) shows a decline.

Himachal Pradesh has had generally excellent levels, though there was a spike in WQI levels in 2011 and 2012. This implies that there is a possibility of groundwater contamination in future.

1. pH levels are improving.
2. Conductivity is generally decreasing, indicating improvement in salinity levels.
3. BOD is showing depleting trends, though there has been a spike in recent years (2013 and 2014).
4. Nitrate and nitrite levels has been going upward, though there was a fall in 2014.
5. Fecal coliform levels generally going upward, though it witnessed a fall in 2013 and 2014.

Kerala has had consistently poor water quality levels over the years and forecasts show further depletion of water quality.

1. Generally low pH, implying acidity. However levels have improved in recent years.
2. Conductivity has generally been decreasing, and predictions depict even more reduction in values.
3. BOD has seen a sudden spike in 2014, implying possibility of bacterial contamination.
4. Nitrate and nitrite levels are have been irregular with occasional spikes. In general, they have been on the bad side.
5. Fecal Coliform are very high, and showing increased growth in the future.

WQI in Maharashtra has been steadily improving, implying improvement in groundwater quality (2015-2017 – 45); though there was a spike in latest data (2014).

1. pH level rapidly increasing, implying increased alkalinity.
2. Conductivity and BOD have been steadily decreasing.
3. Nitrate and nitrite levels have spiked immensely in 2014.
4. Fecal coliform has generally reduced, though it showed a spike in 2014.

Odisha has generally shown excellent WQI; however it displayed a dangerously high WQI for 2014 (possibly due to erroneous data). The WQI values for the forecasted periods are way above the safe limit of 50, due to this possible anomaly.

1. pH, BOD are in acceptable levels.

2. Conductivity has been steadily increasing.
3. Nitrate and nitrite levels have been steadily decreasing, implying improved water quality.
4. Fecal coliform showed a humongous spike in 2014, explaining the abrupt behaviour of WQI. This outlier could possibly be erroneous data.

Rajasthan has excellent WQI, and shows a growing trend.

1. pH is in acceptable levels.
2. Conductivity is high and is continuously growing, implying increase in salinity.
3. BOD is also increasing, implying growth of harmful bacteria.
4. Nitrate and nitrite levels are at acceptable range.
5. Fecal coliform levels have been gradually reducing.

Uttar Pradesh displays very steep increase in WQI since 2011 (2014 – 66, 2015-2017 – 63)

1. pH is steadily increasing, implying alkalinity.
2. Conductivity shows irregular trends, though they generally have been reducing.
3. BOD saw a huge spike in 2013, though it got restored in 2014. However the possibility of bacteria can't be ignored.
4. Nitrate and nitrite levels show irregular trends, but they are rather high.
5. Fecal coliform levels are also growing tremendously.

WQI in West Bengal has been steadily increasing, implying poor quality. Predictions also show further deterioration in quality.

1. pH is generally in acceptable levels, though it is quickly moving towards alkalinity.
2. Conductivity is falling down, implying decreased salinity.
3. BOD and Nitrate and Nitrite levels are at acceptable levels, though they show growing trends.
4. Fecal coliform shows high rate of growth, implying contamination in the future.

Conclusions

Groundwater quality is deteriorating at an alarming rate owing to various factors. Most states are facing issues of alkalinity, salinity and possible bacterial contamination. The analysis reveals that the groundwater of the affected states need some degree of treatment before consumption and it also needs to be protected from the hazards of further contamination. Therefore it is highly critical to take into consideration the remedial measures for the same, else groundwater could get polluted to a great extent in these states.

Scope of Further Work

1. Analyze other states which are currently in safe zone, to identify potential ones whose groundwater could get contaminated in future.
2. Drill down into possible erroneous data and find out causes of the error. This would probably require physical visits to the sites, as well as meetings with the data collection team.
3. Analyze the various factors that are contributing to the fluctuations of the various parameters. This would include gathering other data, possibly not related to the above data.
4. Improve the accuracy of the predictions by collecting more accurate data and adding more parameters to the model.
5. Develop a regression model to understand the interaction of the parameters and come up with ways to control the increase in WQI Index levels.
6. Study can also be extended to surface water (ponds, lakes, creeks, etc).

Acknowledgements

The satiation and euphoria that accompany the successful completion of the project would be incomplete without the mention of the people who made it possible.

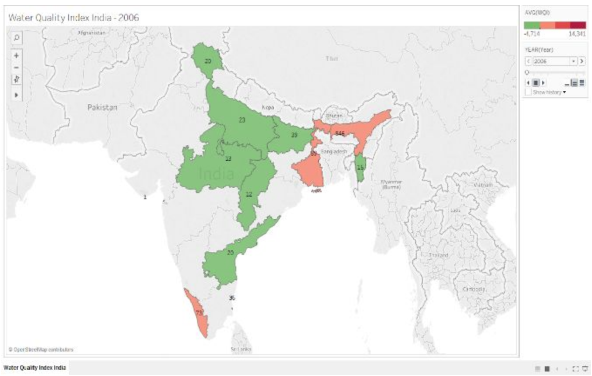
We would like to take the opportunity to thank and express our deep sense of gratitude to our course mentor Prof. Arnab Kumar Laha and our project mentor Prof. Vineet Virmani. We are greatly indebted to both of them for providing their valuable guidance at all stages of the study, their advice, constructive suggestions, positive and supportive attitude and continuous encouragement, without which it would have not been possible to complete the project. We would also like to thank all our other mentors (Prof. Goutam Dutta, Prof. Apratim Guha, Prof. Sanjay Verma, Prof. Vineet Virmani, Prof. Karthik Sriram, Prof. Kavitha Ranganathan, Prof. Ankur Sinha) whose lectures and guidance have immensely helped us to complete our project.

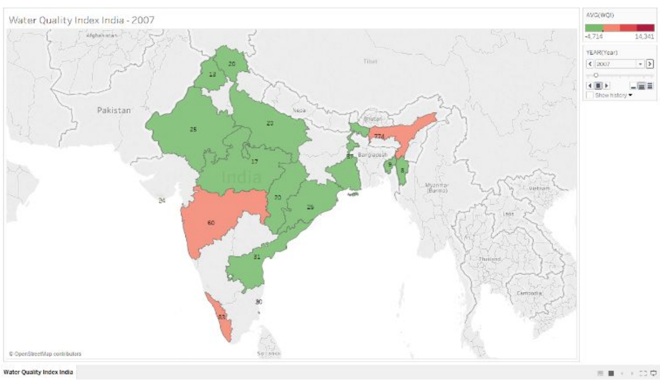
We would also like to thank the team at Hughes, whose infrastructure and support immensely helped us and allowed us to complete the course on a virtual platform, despite us being physically absent from the college campus.

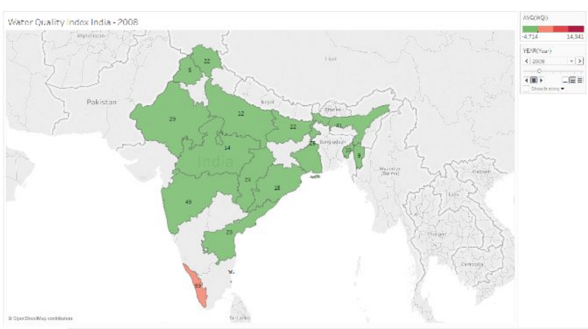
Last but not the least, we would like to thank the Government of India and the various government officials who gathered the data and made it available on public domain, thus allowing us to work on the same.

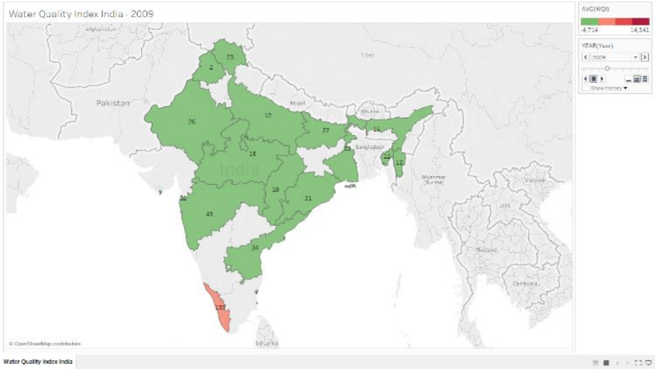
Annexure

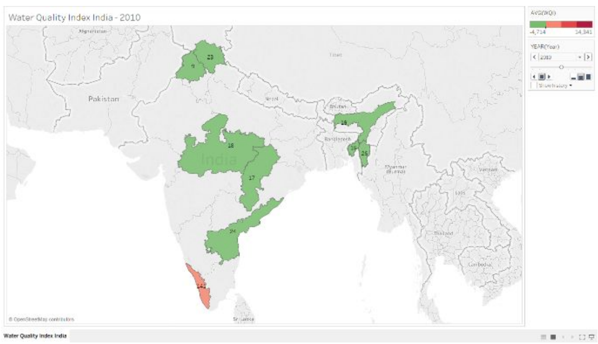
Tableau Plots

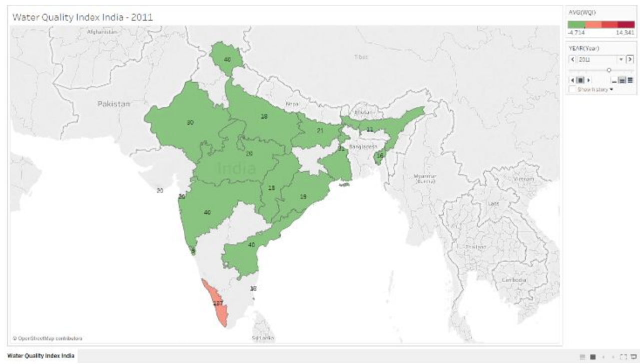


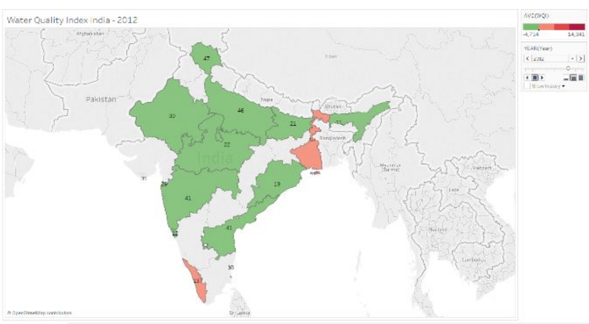


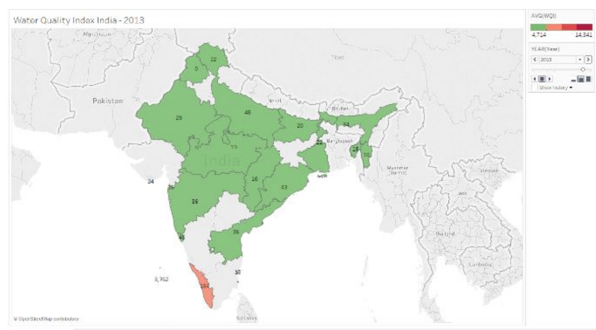


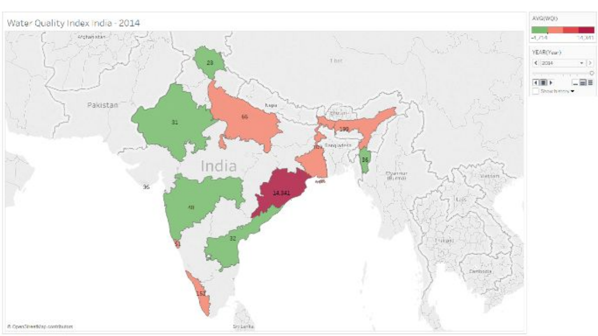

Water Quality Index India - 2009
Pakistan
India
-4,714
14,341
Water Quality Index India

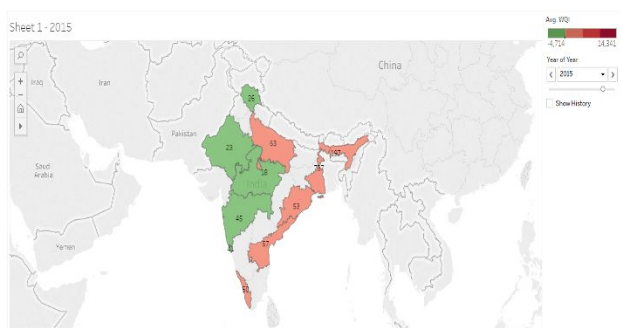

Water Quality Index India - 2010
Pakistan
India
-4,714
14,341
Water Quality Index India

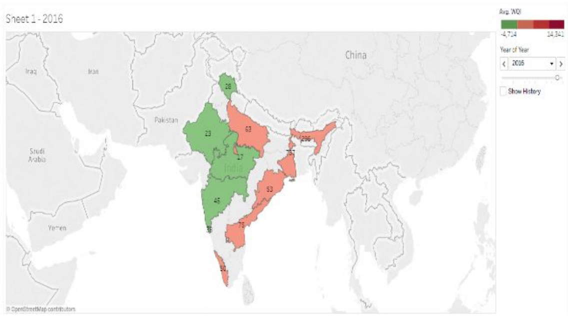

Water Quality Index India - 2011
Pakistan
India
-4,714
14,341
Water Quality Index India


Water Quality Index India - 2012
Pakistan
India
-4,714
14,341
Water Quality Index India


Water Quality Index India - 2013
Pakistan
India
-4,714
14,341
Water Quality Index India


Water Quality Index India - 2014
Pakistan
India
14,341
Water Quality Index India


Sheet 1 - 2015
Avg. WQI
-4,714
14,341
Year of Year
2015
Show History
China
Pakistan
Iran
Saudi Arabia
Yemen
India
© OpenStreetMap contributors



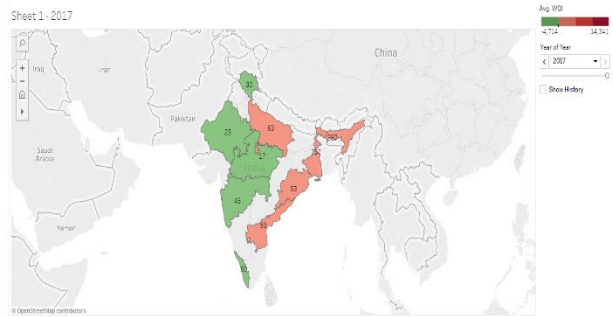
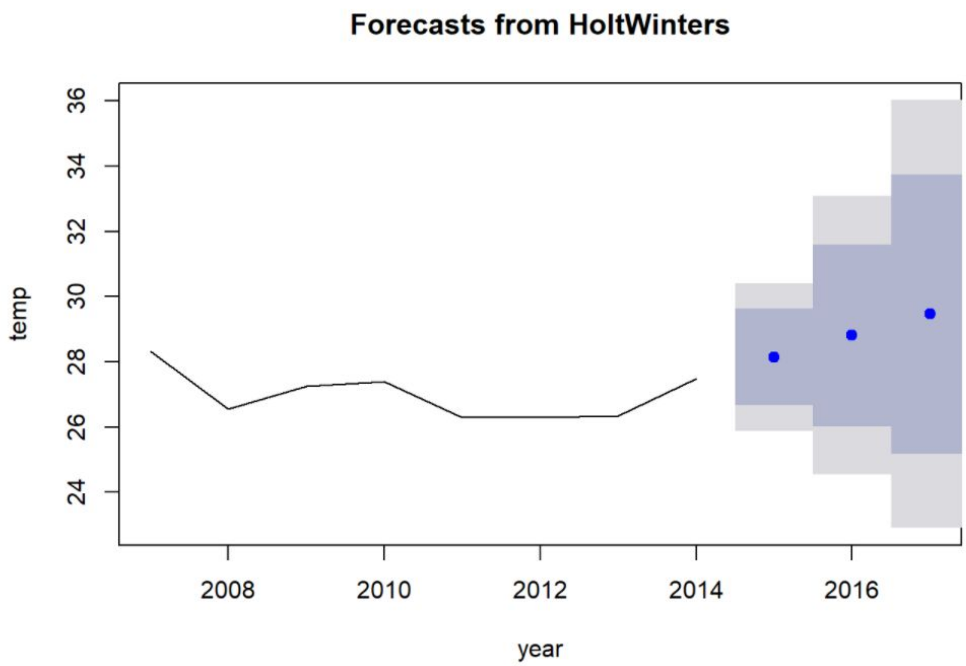


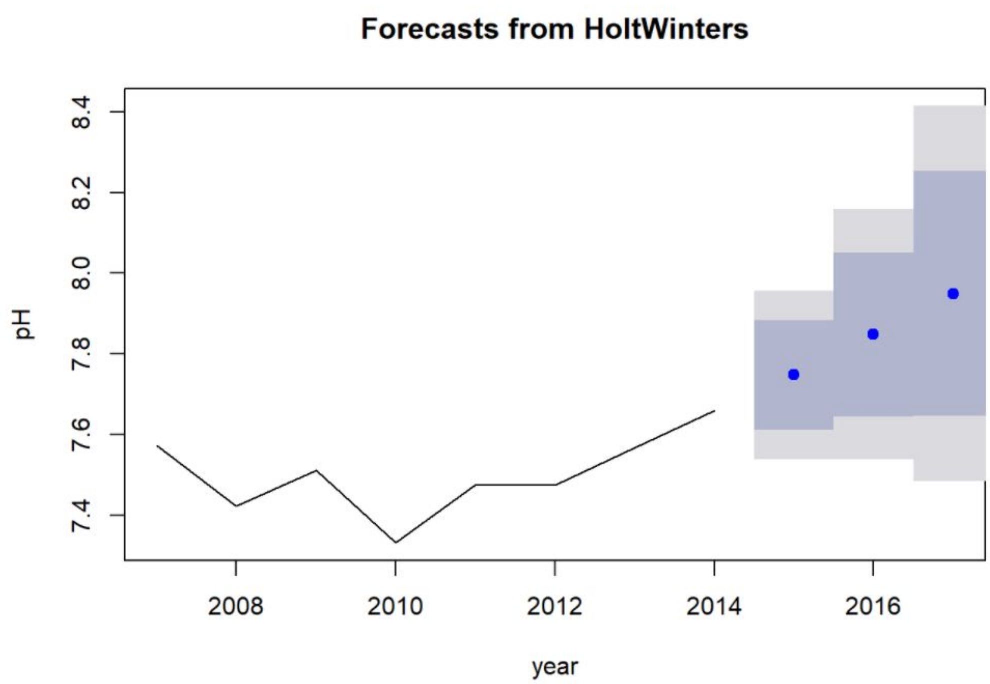
Tableau plots:
https://public.tableau.com/profile/shoumik.goswami#!/

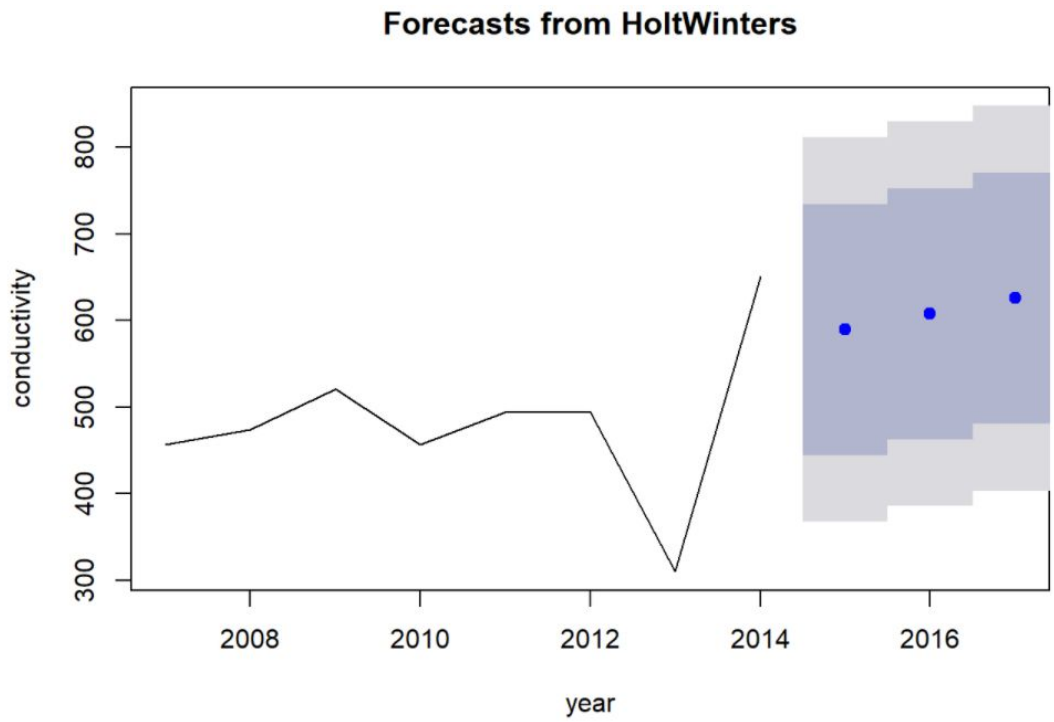
Graphs

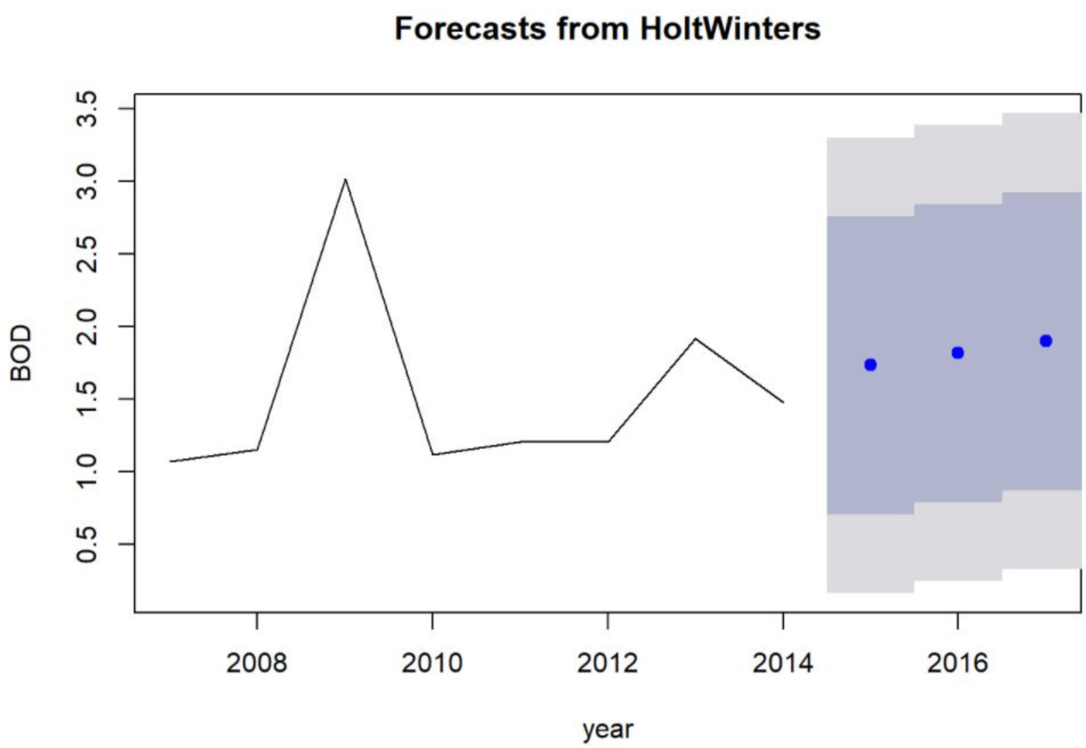
https://github.com/VDSouz/IndiaWaterQuality/tree/master/States

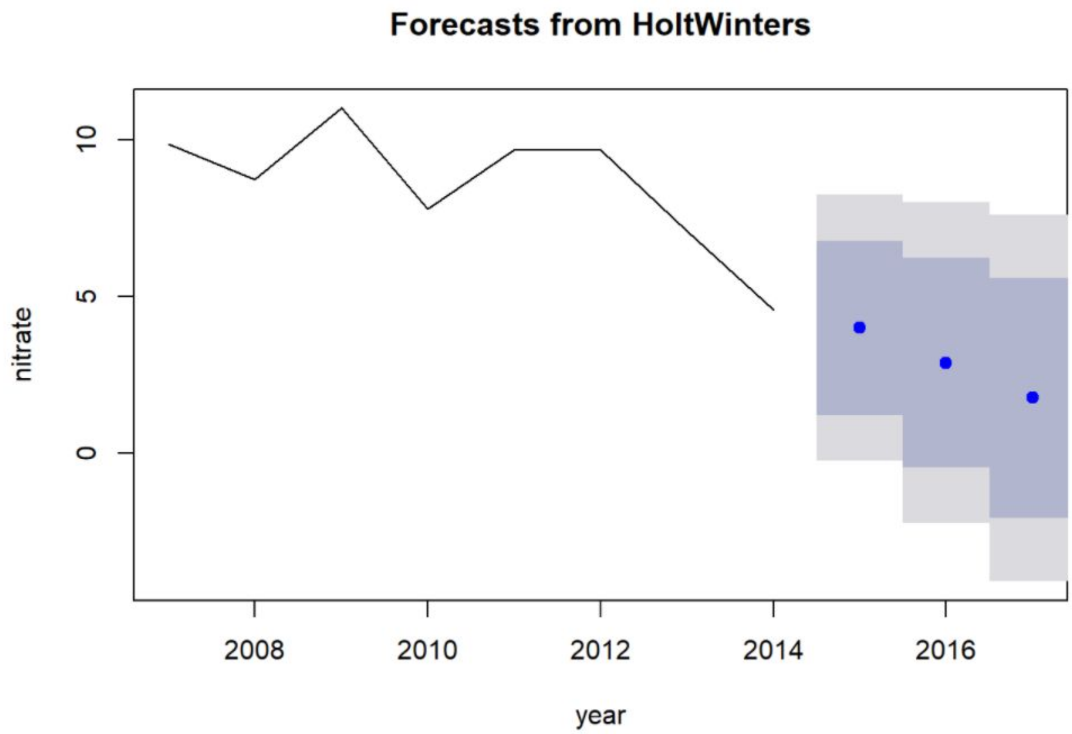
Andhra Pradesh



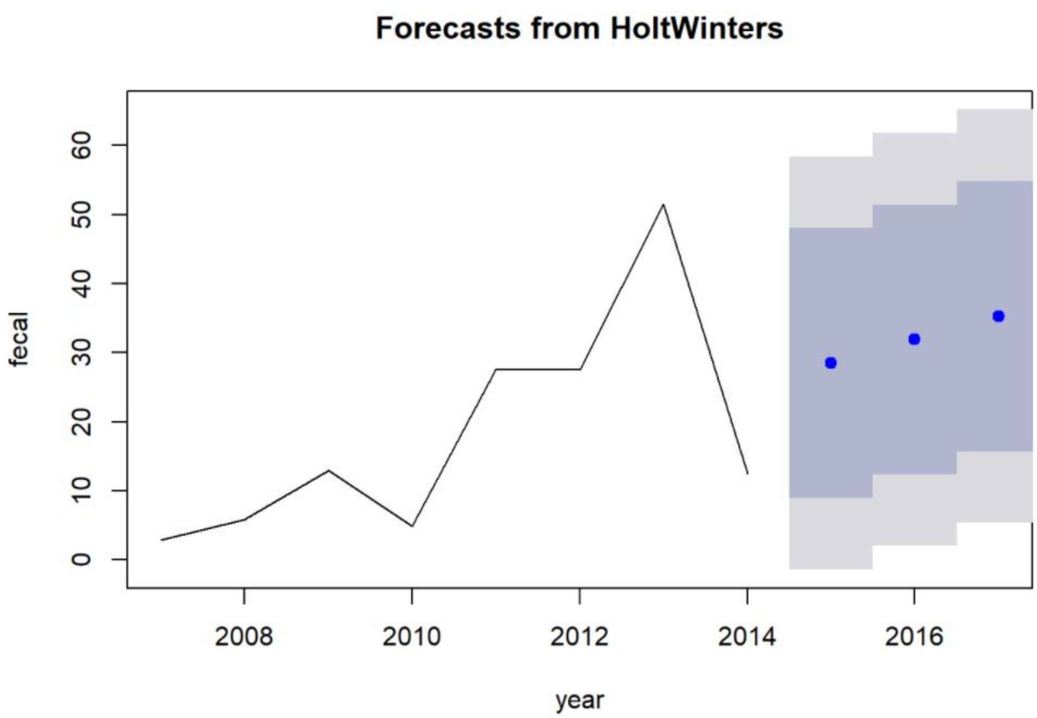



Forecasts from HoltWinters
conductivity
300
400
500
600
700
800
2008
2010
2012
2014
2016
year


Forecasts from HoltWinters
BOD
0.5
1.0
1.5
2.0
2.5
3.0
3.5
2008
2010
2012
2014
2016
year

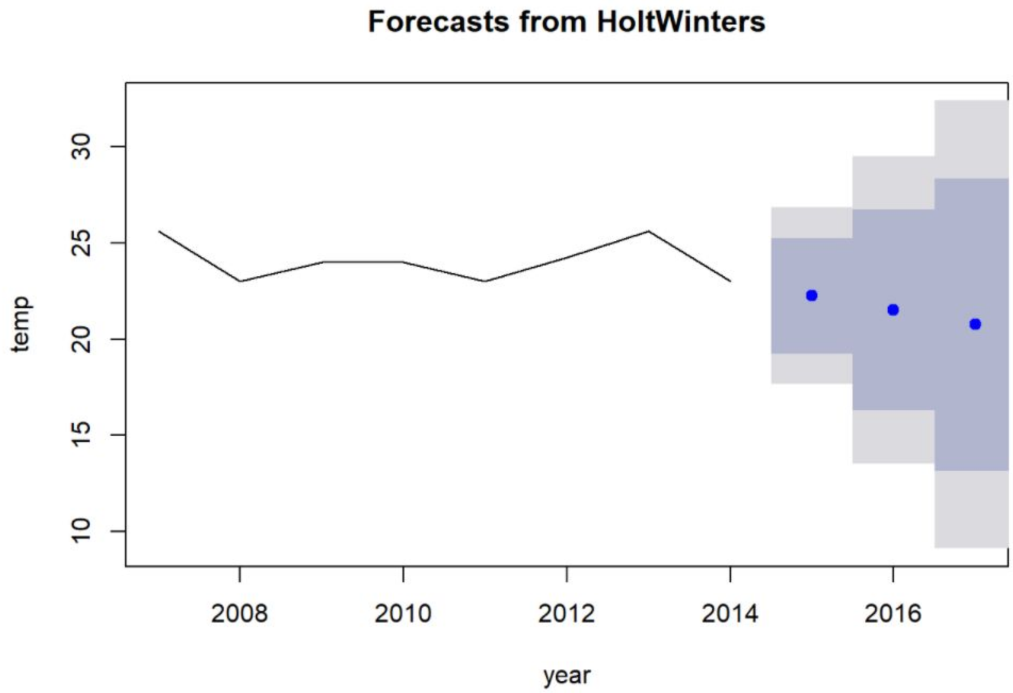


Forecasts from HoltWinters
nitrate
10
5
0
2008
2010
2012
2014
2016
year

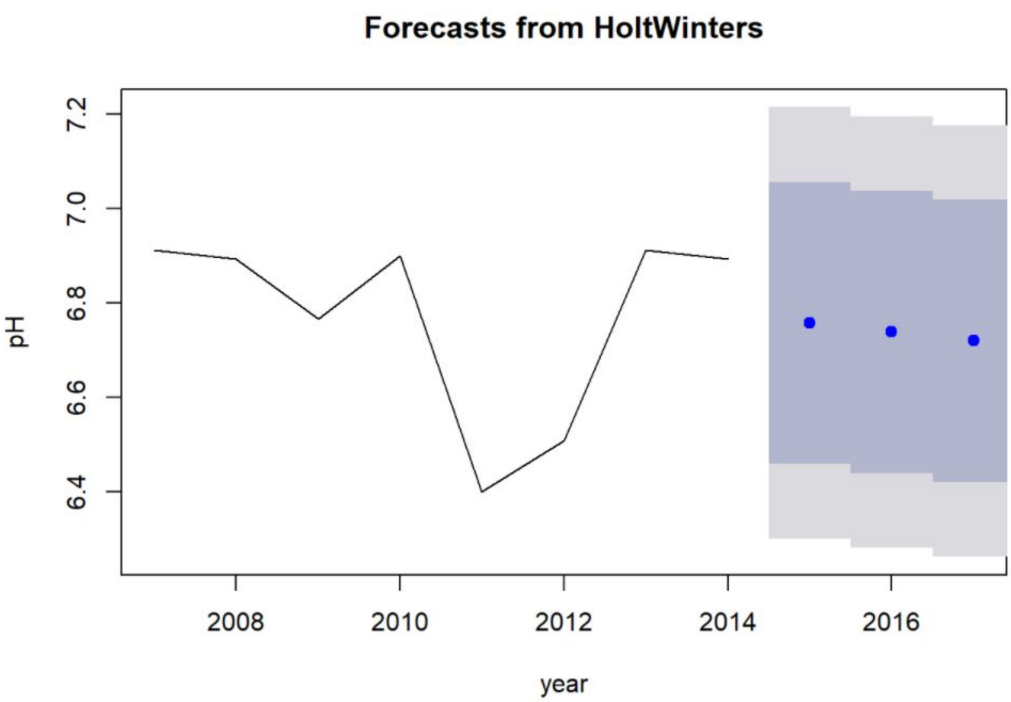


Forecasts from HoltWinters
fecal
60
50
40
30
20
10
0
2008
2010
2012
2014
2016
year

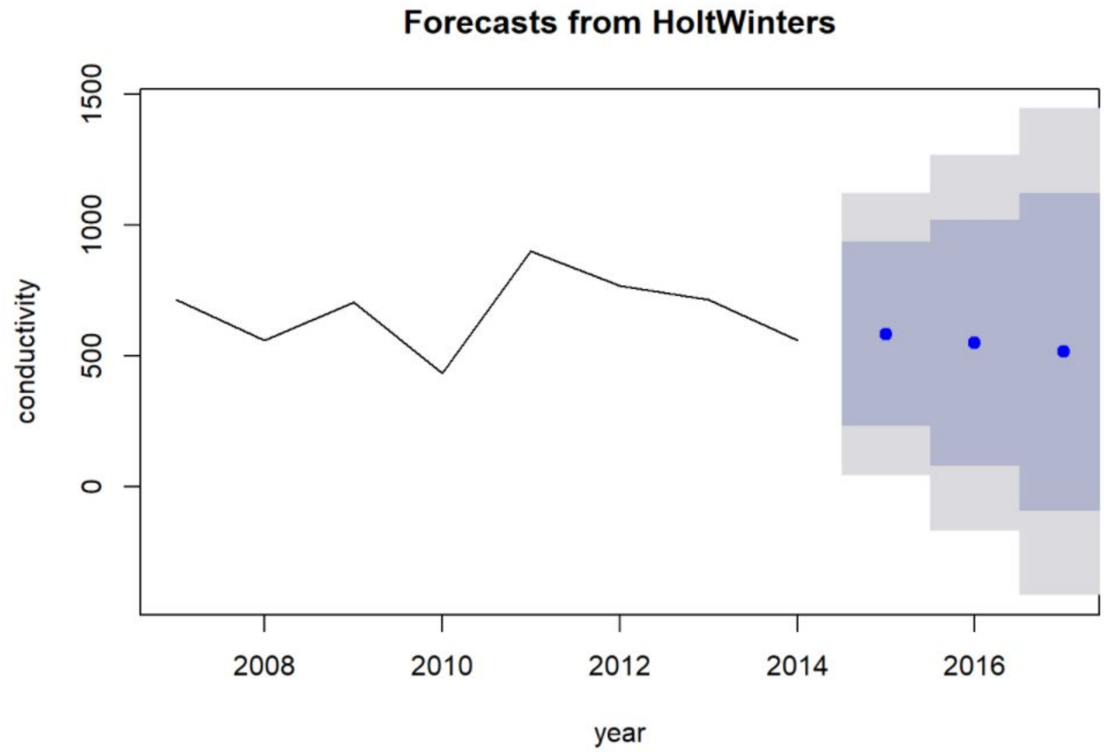
Assam

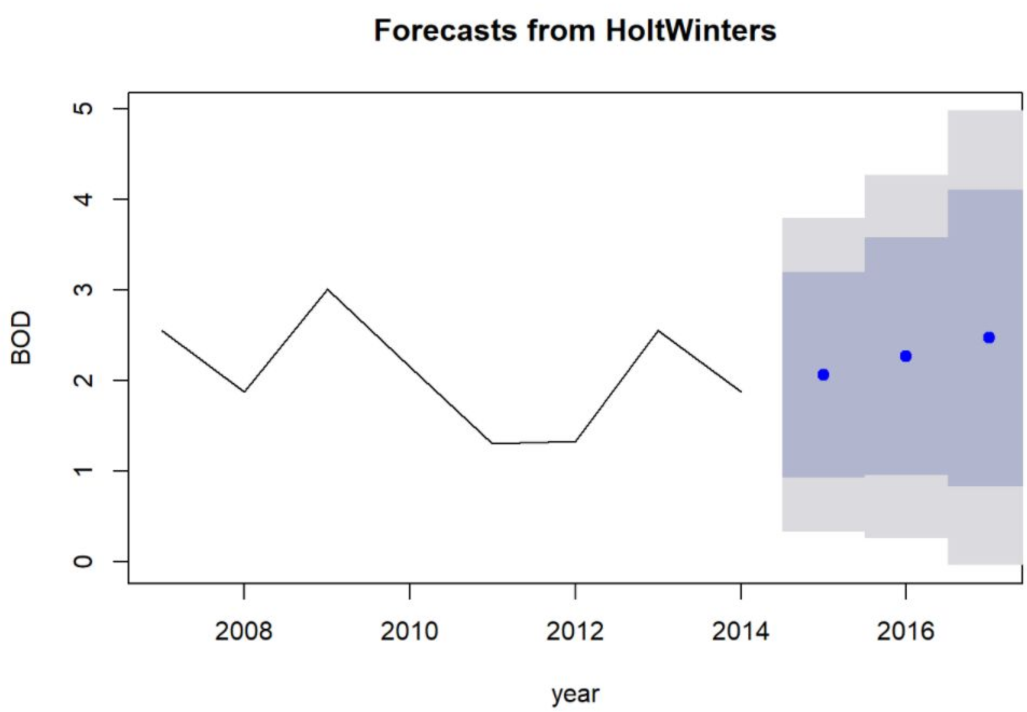


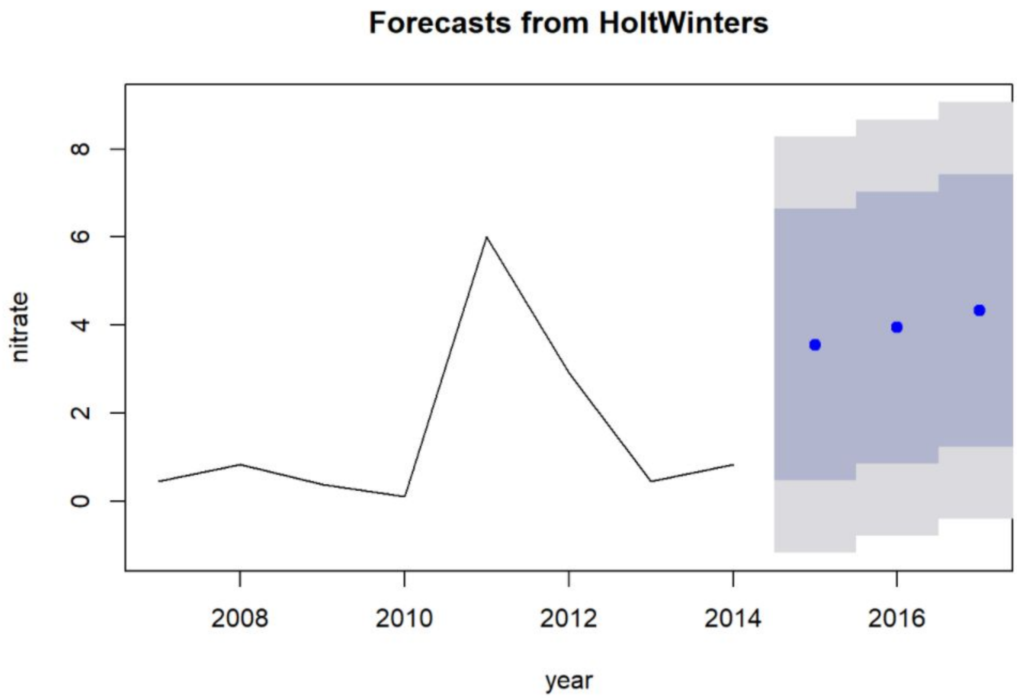
Forecasts from HoltWinters
temp
10
15
20
25
30
2008
2010
2012
2014
2016
year

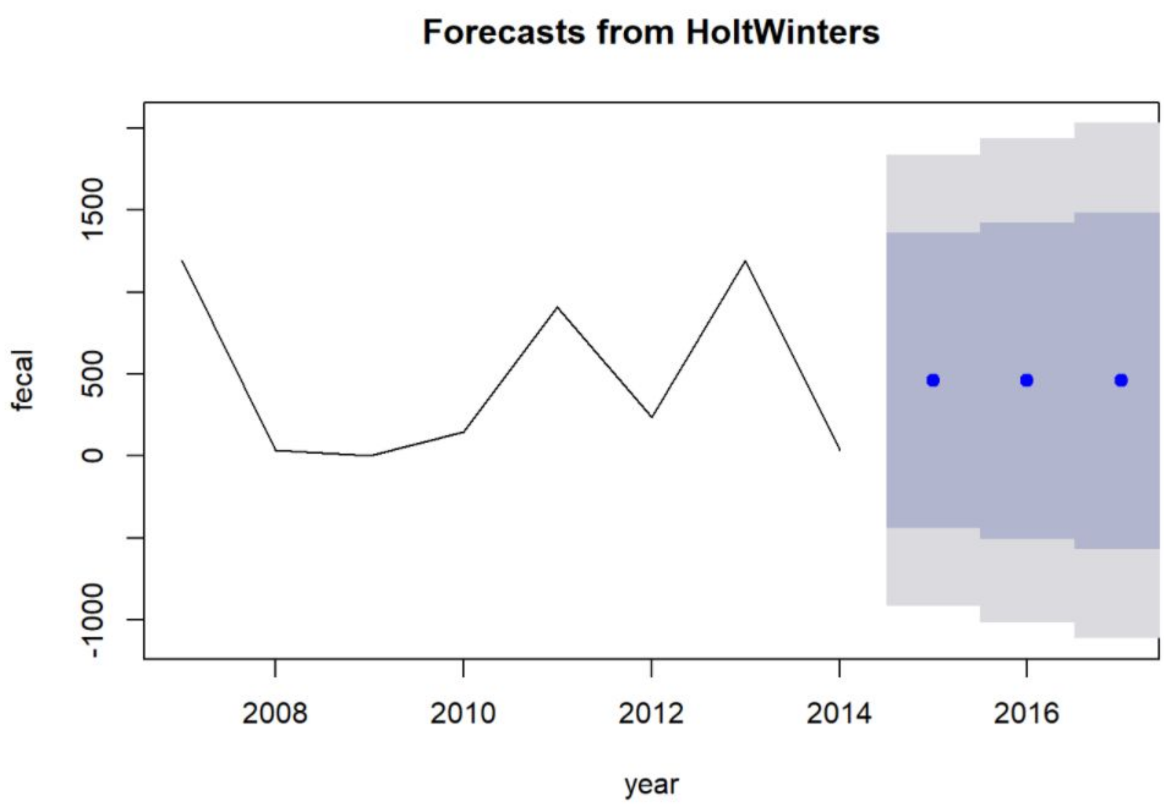


Forecasts from HoltWinters
pH
6.4
6.6
6.8
7.0
7.2
2008
2010
2012
2014
2016
year

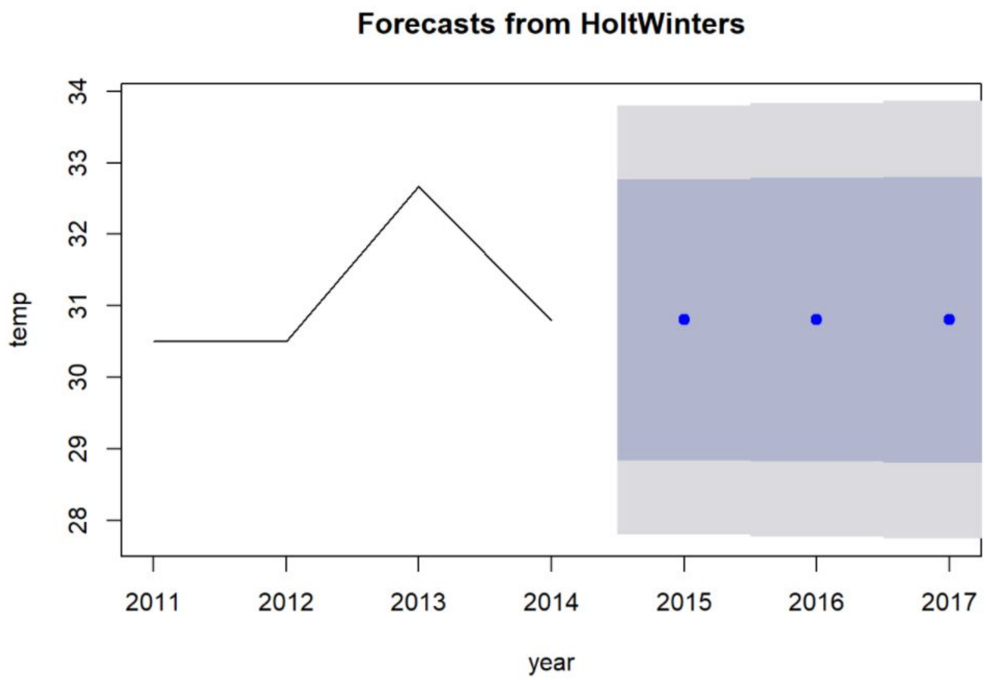

Forecasts from HoltWinters
conductivity
year
0
500
1000
1500
2008
2010
2012
2014
2016


Forecasts from HoltWinters
BOD
year
0
1
2
3
4
5
2008
2010
2012
2014
2016

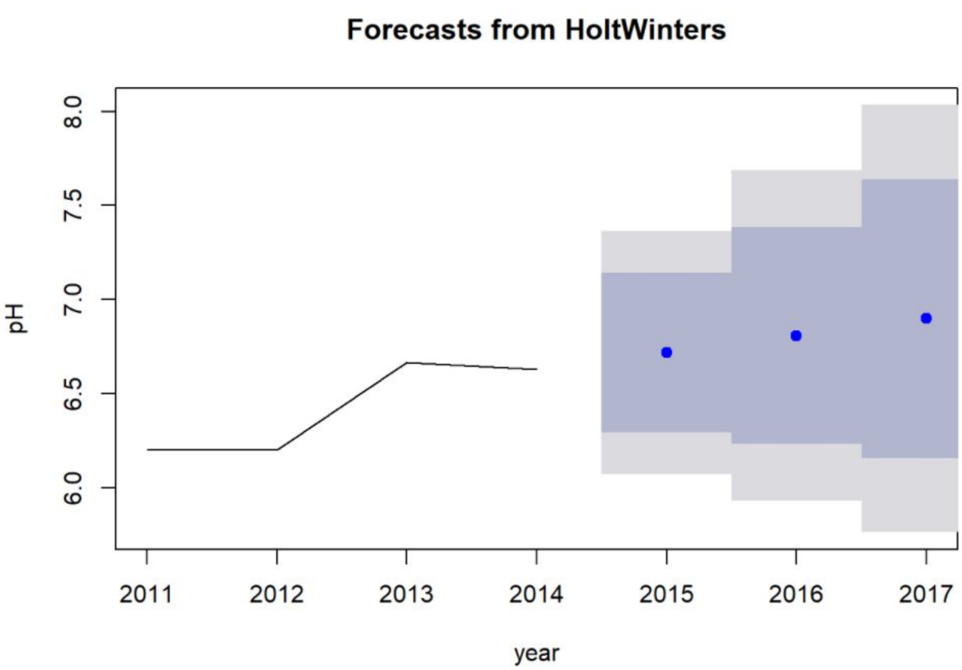

Forecasts from HoltWinters
nitrate
0
2
4
6
8
2008
2010
2012
2014
2016
year


Forecasts from HoltWinters
fecal
-1000
0
500
1500
2008
2010
2012
2014
2016
year

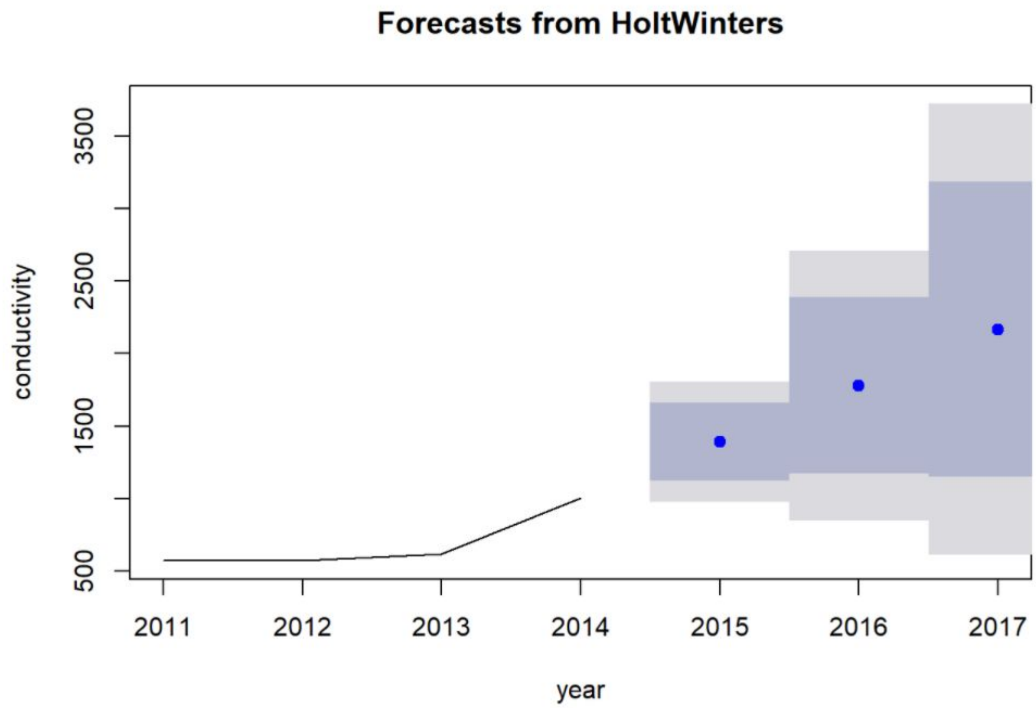
Goa

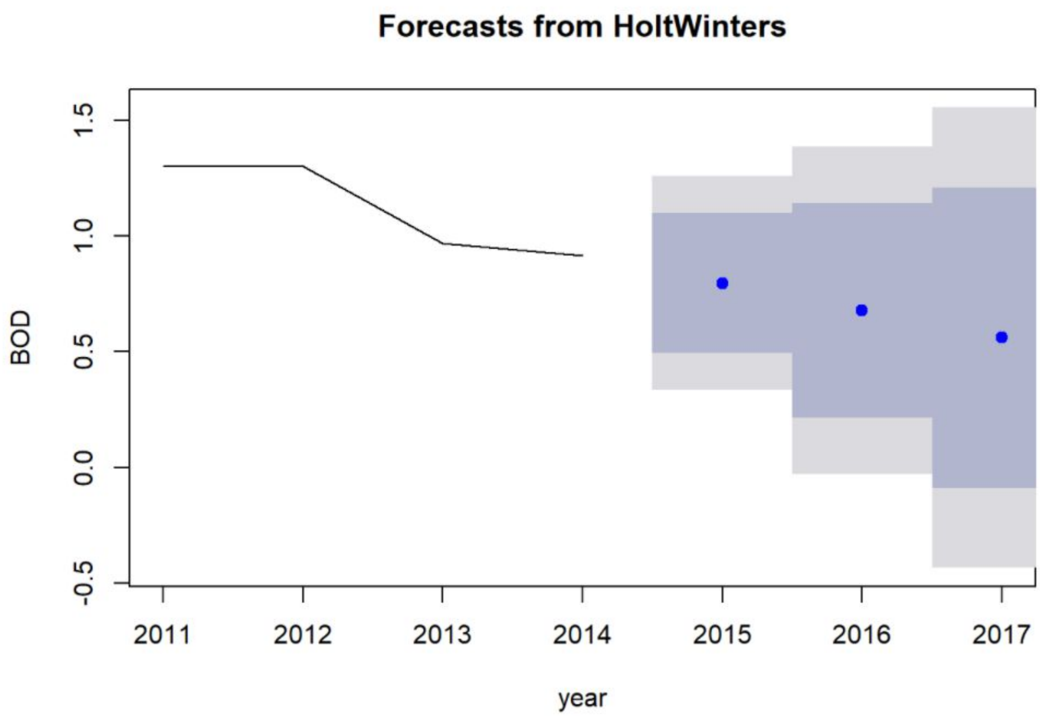


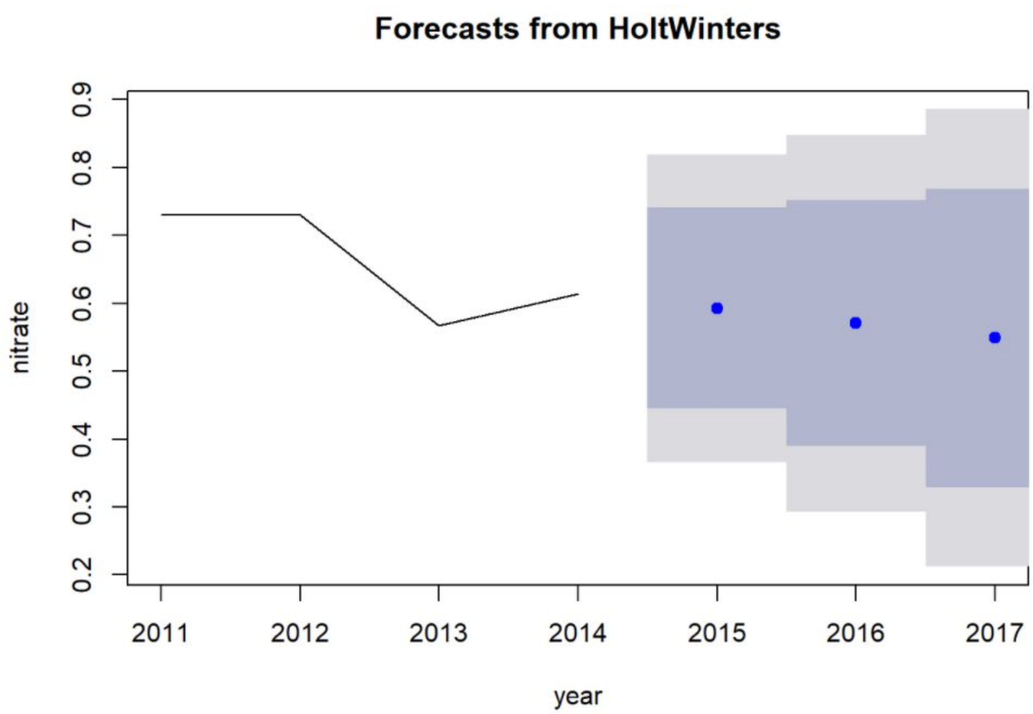
Forecasts from HoltWinters
temp
34
33
32
31
30
29
28
2011
2012
2013
2014
2015
2016
2017
year

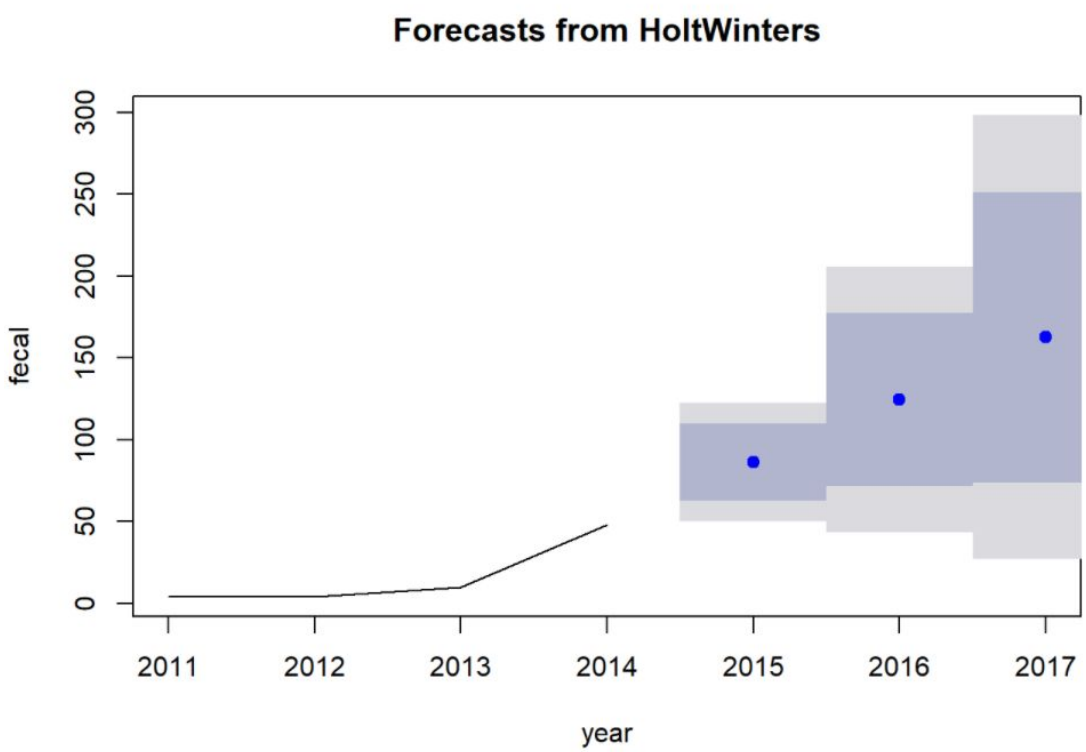


Forecasts from HoltWinters
pH
8.0
7.5
7.0
6.5
6.0
2011
2012
2013
2014
2015
2016
2017
year

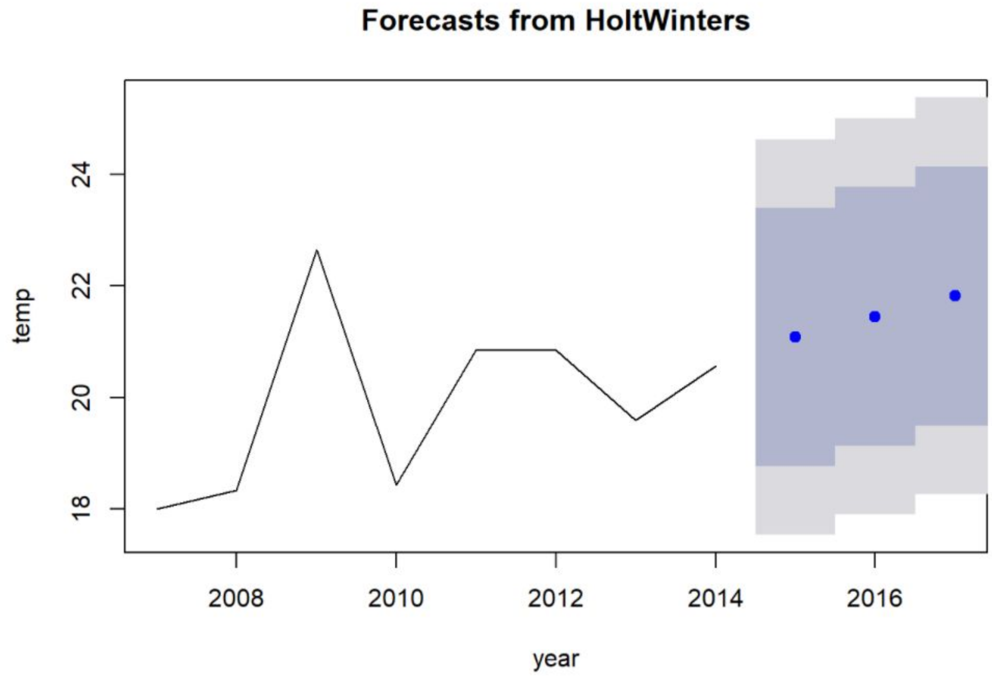

Forecasts from HoltWinters
conductivity
500
1500
2500
3500
2011
2012
2013
2014
2015
2016
2017
year

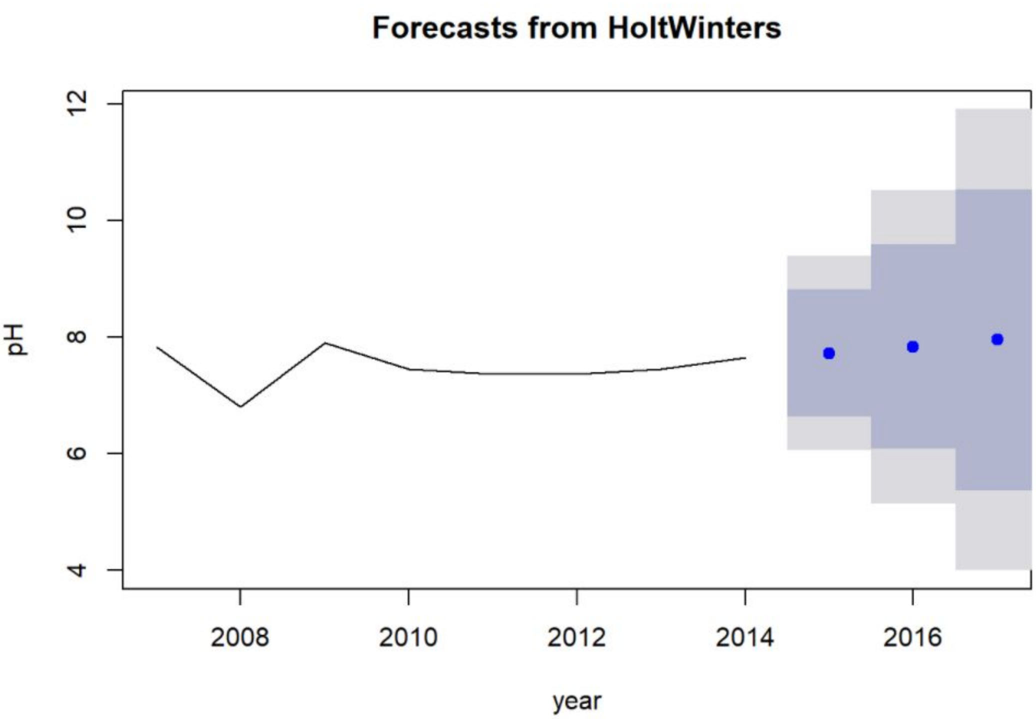

Forecasts from HoltWinters
BOD
-0.5
0.0
0.5
1.0
1.5
2011
2012
2013
2014
2015
2016
2017
year

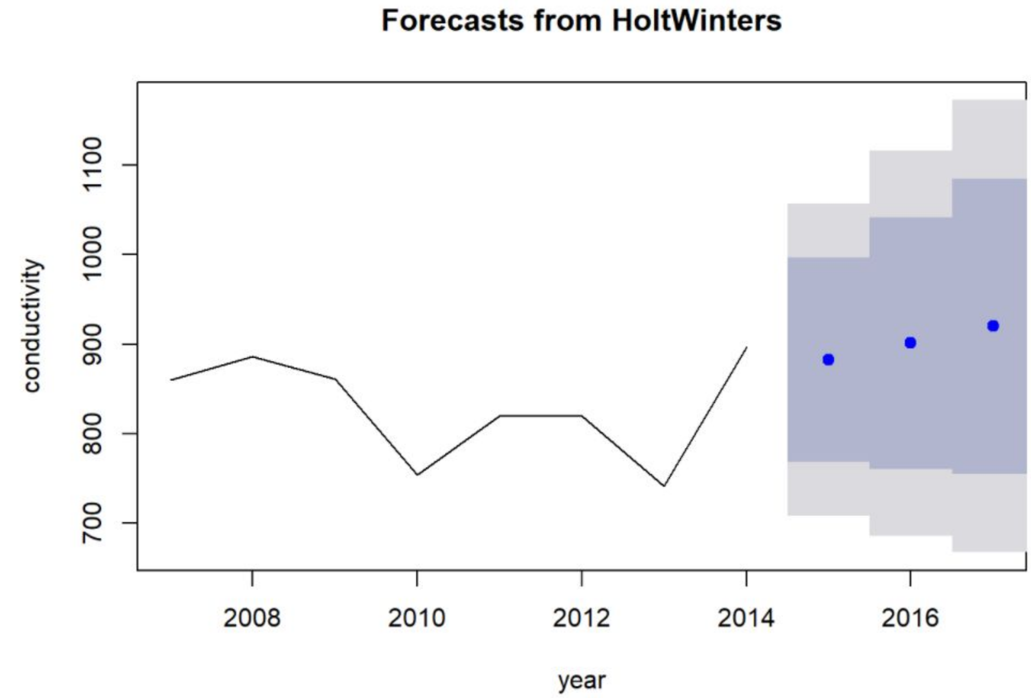

Forecasts from HoltWinters
nitrate
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
2011
2012
2013
2014
2015
2016
2017
year

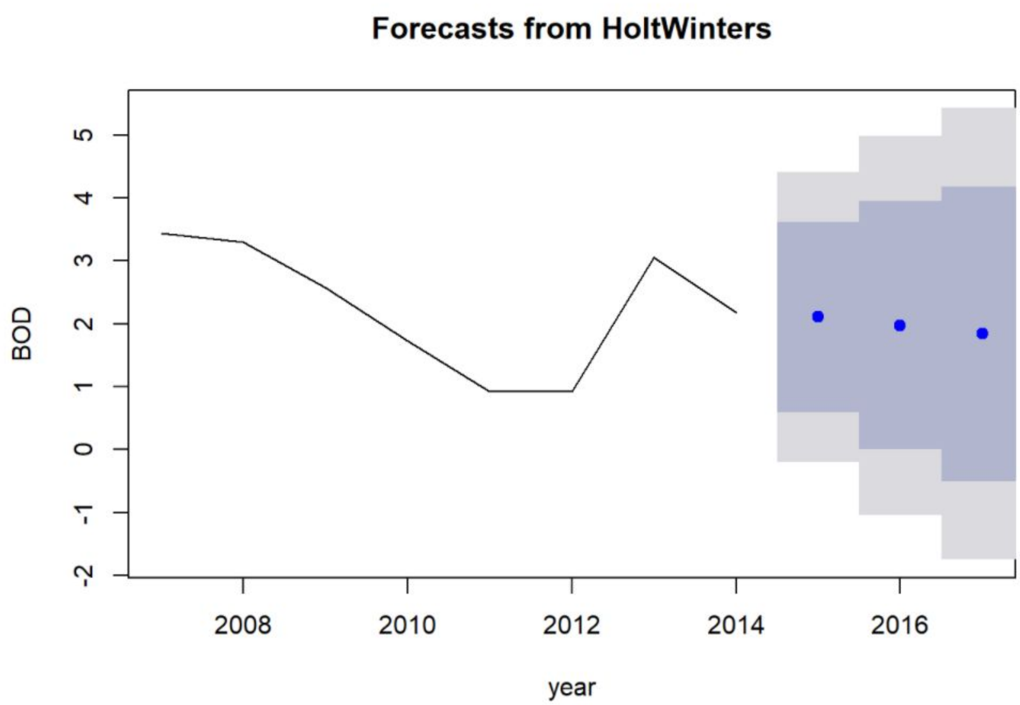

Forecasts from HoltWinters
fecal
0
50
100
150
200
250
300
2011
2012
2013
2014
2015
2016
2017
year

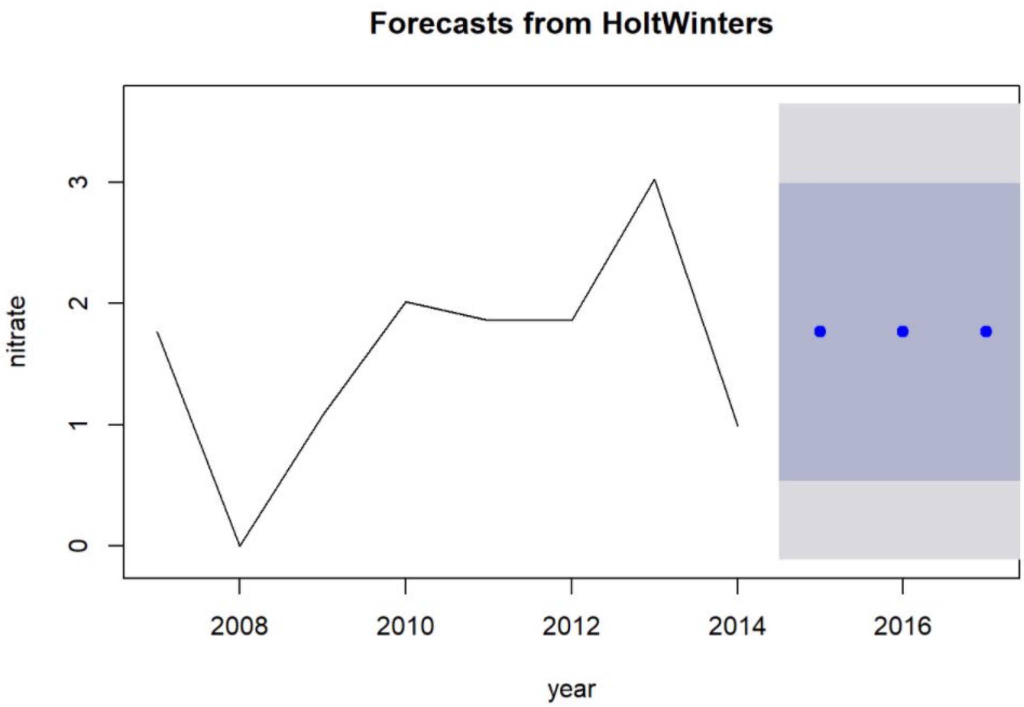
Himachal Pradesh

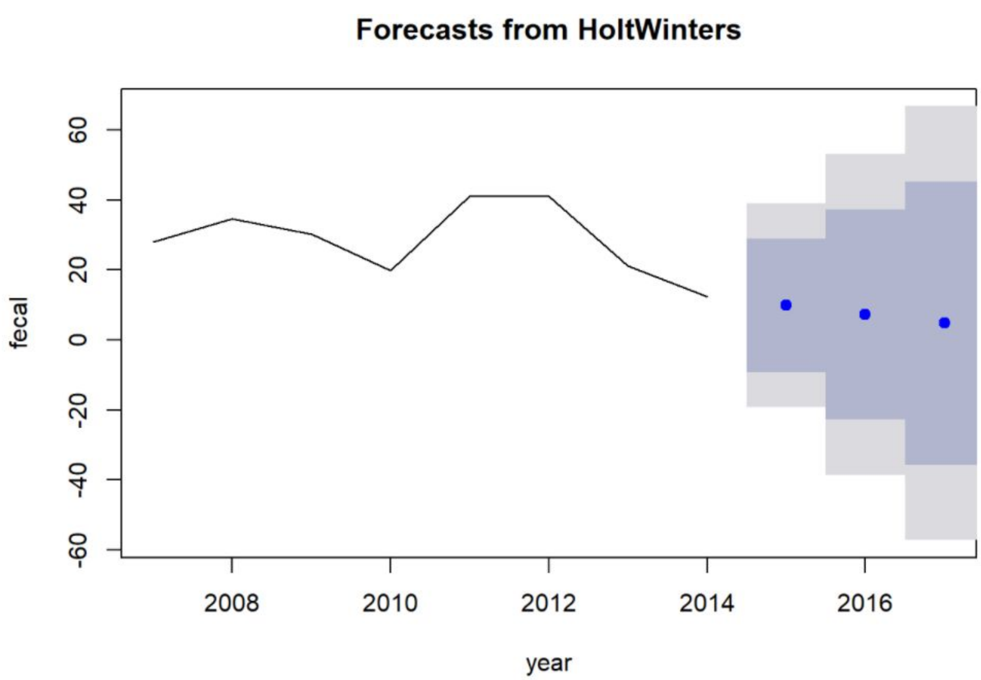




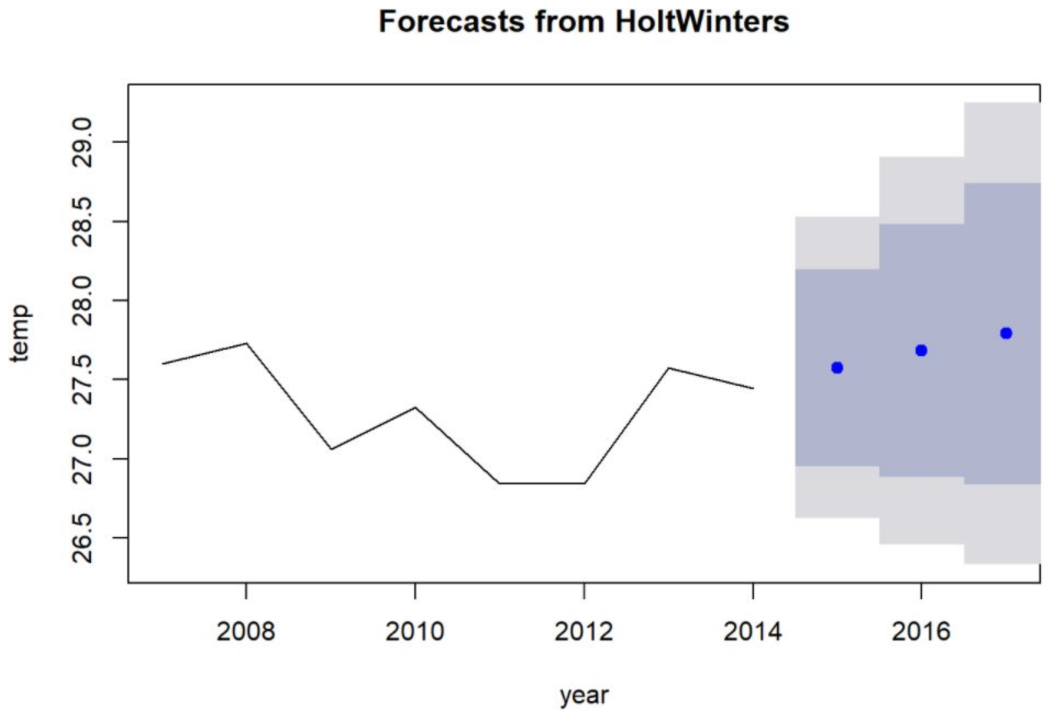

Forecasts from HoltWinters
conductivity
700
800
900
1000
1100
2008
2010
2012
2014
2016
year


Forecasts from HoltWinters
BOD
-2
-1
0
1
2
3
4
5
2008
2010
2012
2014
2016
year

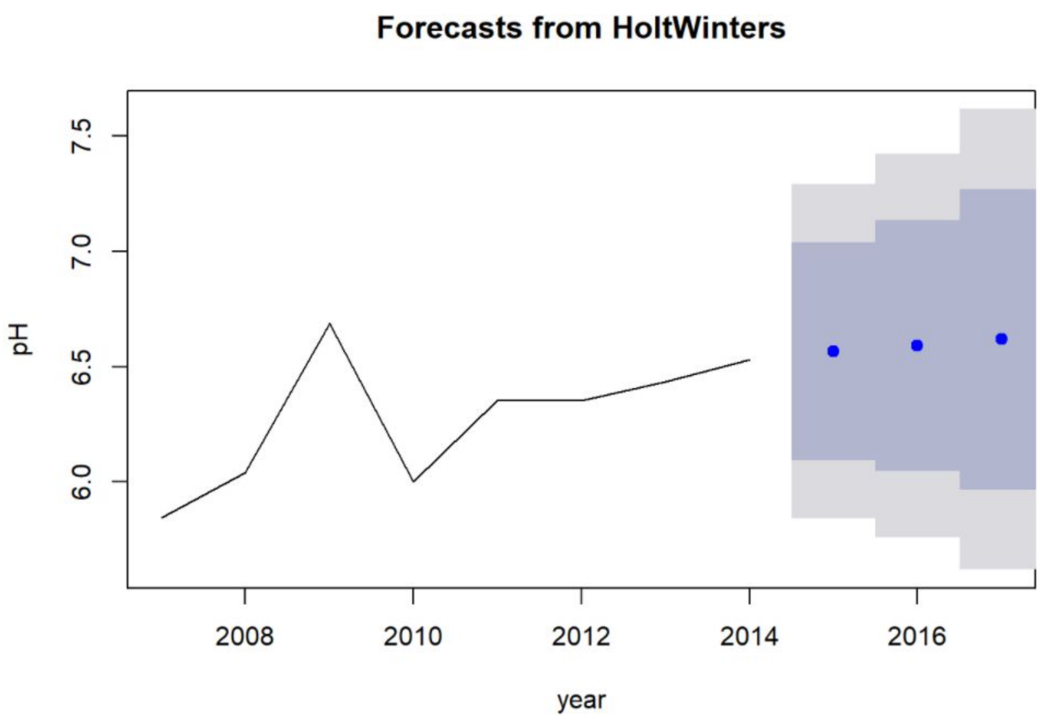

Forecasts from HoltWinters
nitrate
0
1
2
3
2008
2010
2012
2014
2016
year


Forecasts from HoltWinters
fecal
-60
-40
-20
0
20
40
60
2008
2010
2012
2014
2016
year

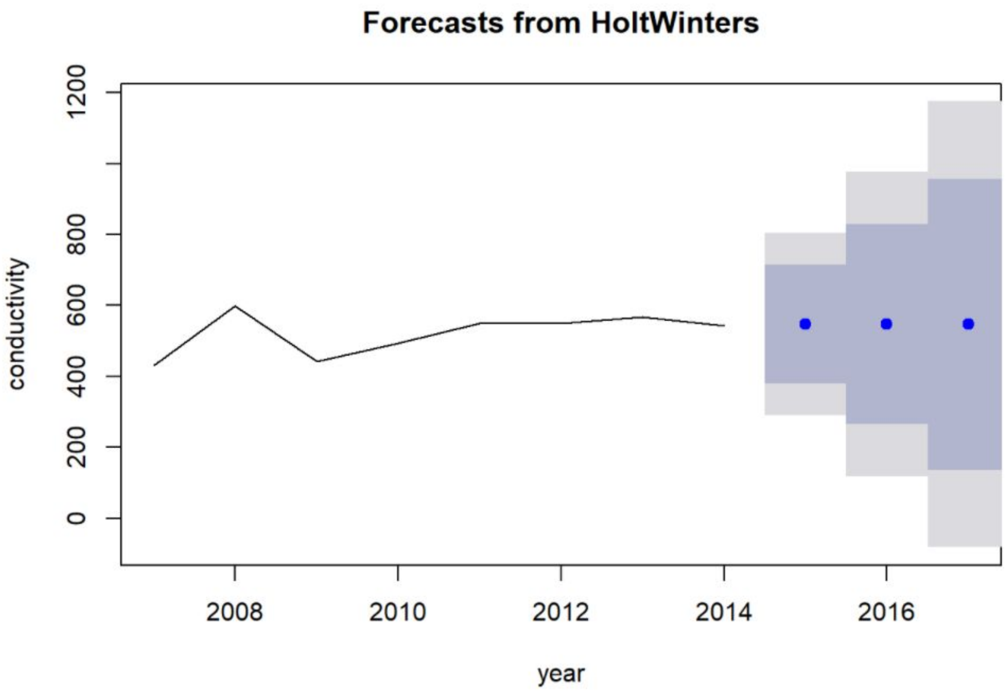
Kerala

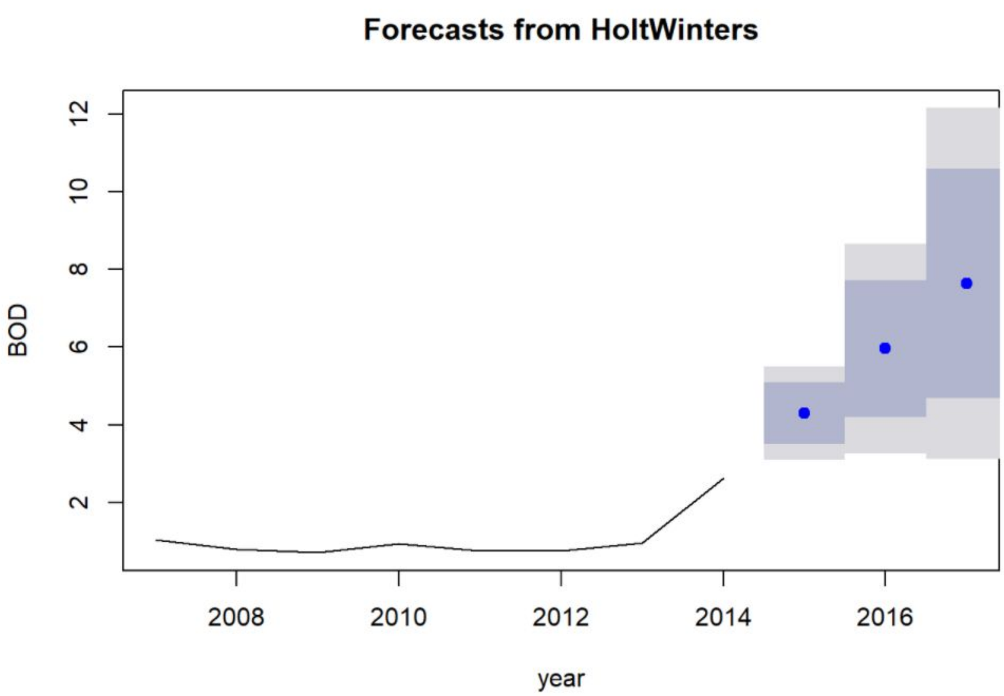


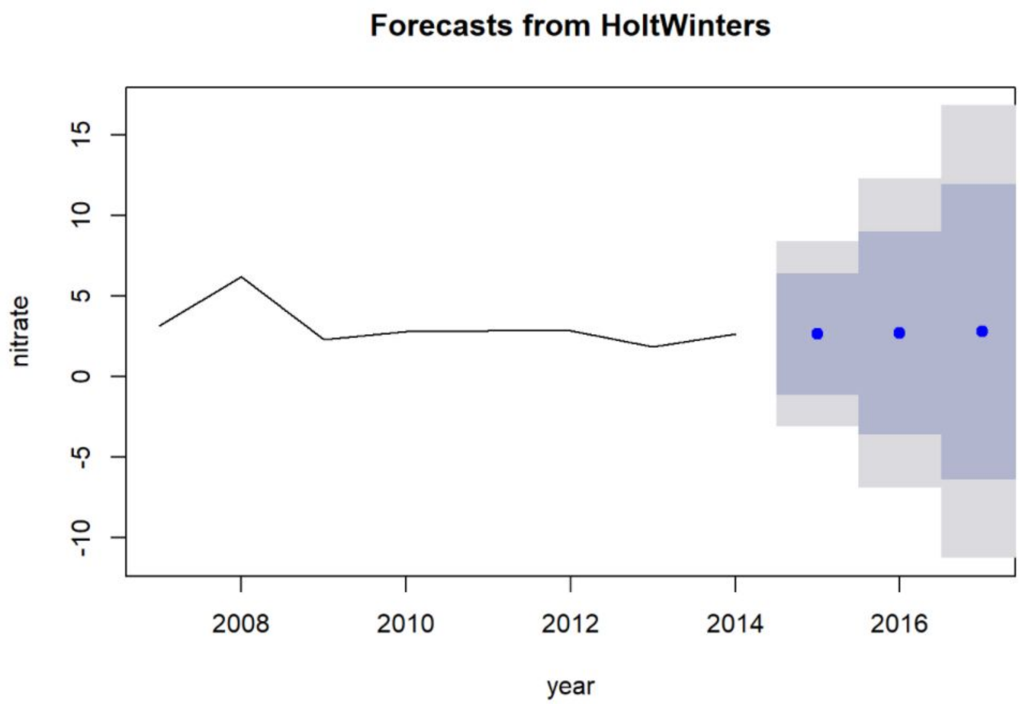
Forecasts from HoltWinters
temp
29.0
28.5
28.0
27.5
27.0
26.5
2008
2010
2012
2014
2016
year

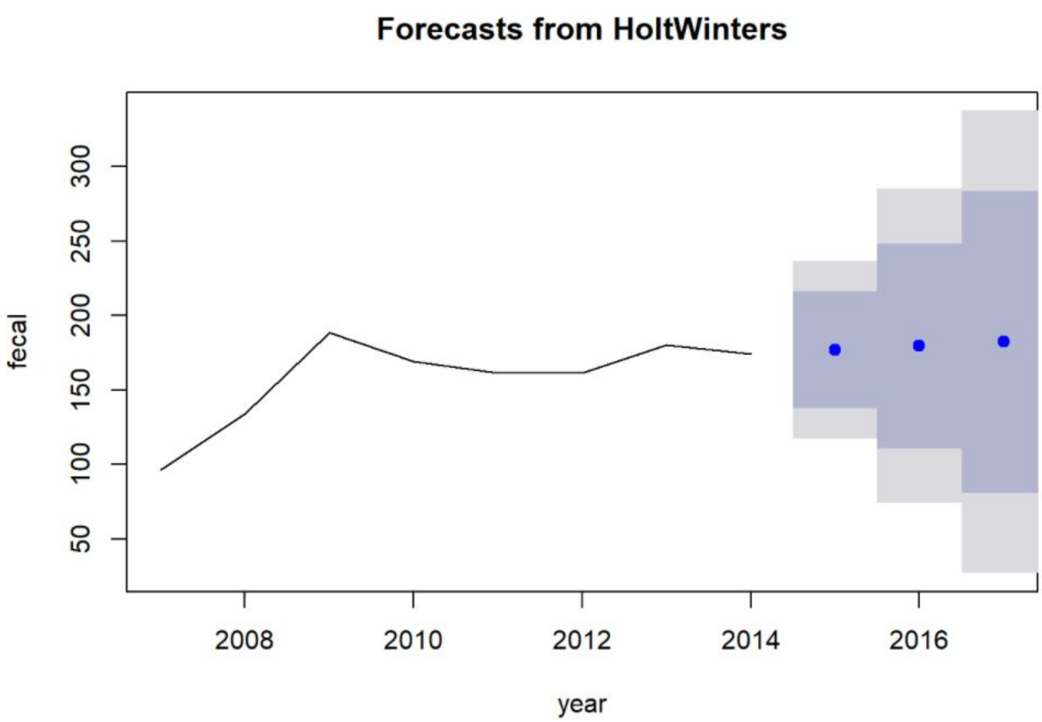


Forecasts from HoltWinters
pH
7.5
7.0
6.5
6.0
2008
2010
2012
2014
2016
year

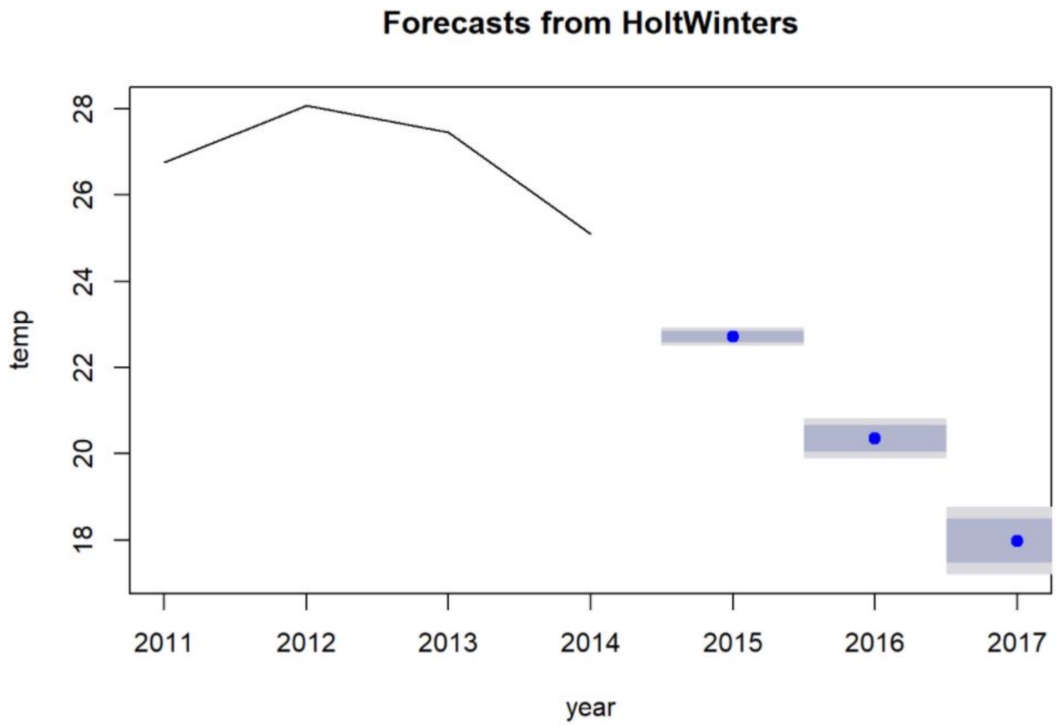

Forecasts from HoltWinters
conductivity
0
200
400
600
800
1200
2008
2010
2012
2014
2016
year

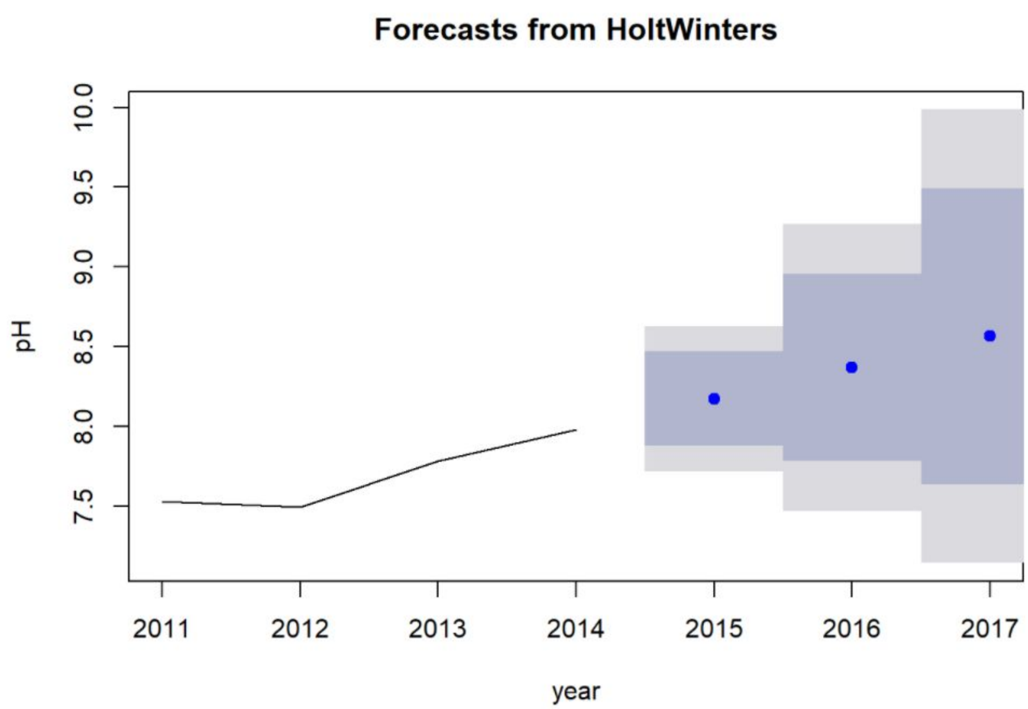

Forecasts from HoltWinters
BOD
2
4
6
8
10
12
2008
2010
2012
2014
2016
year

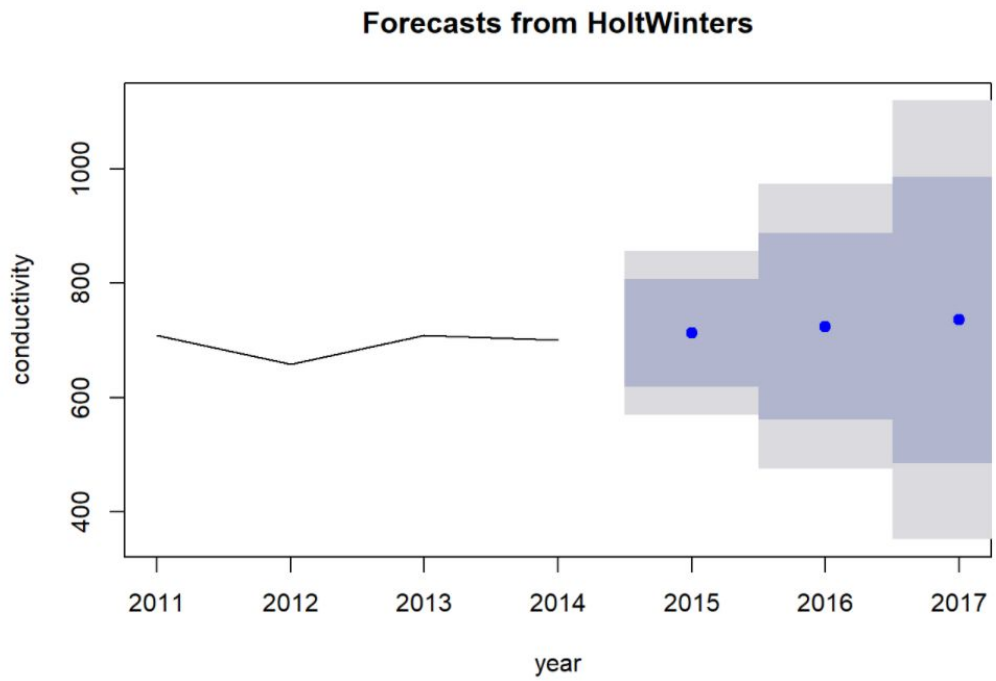

Forecasts from HoltWinters
nitrate
15
10
5
0
-5
-10
2008
2010
2012
2014
2016
year

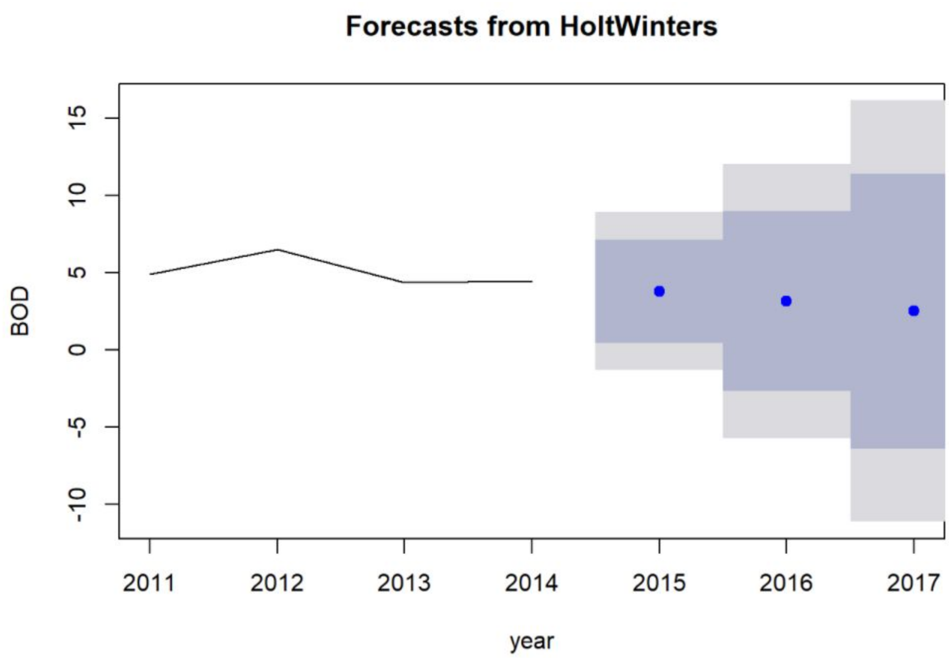

Forecasts from HoltWinters
fecal
300
250
200
150
100
50
2008
2010
2012
2014
2016
year

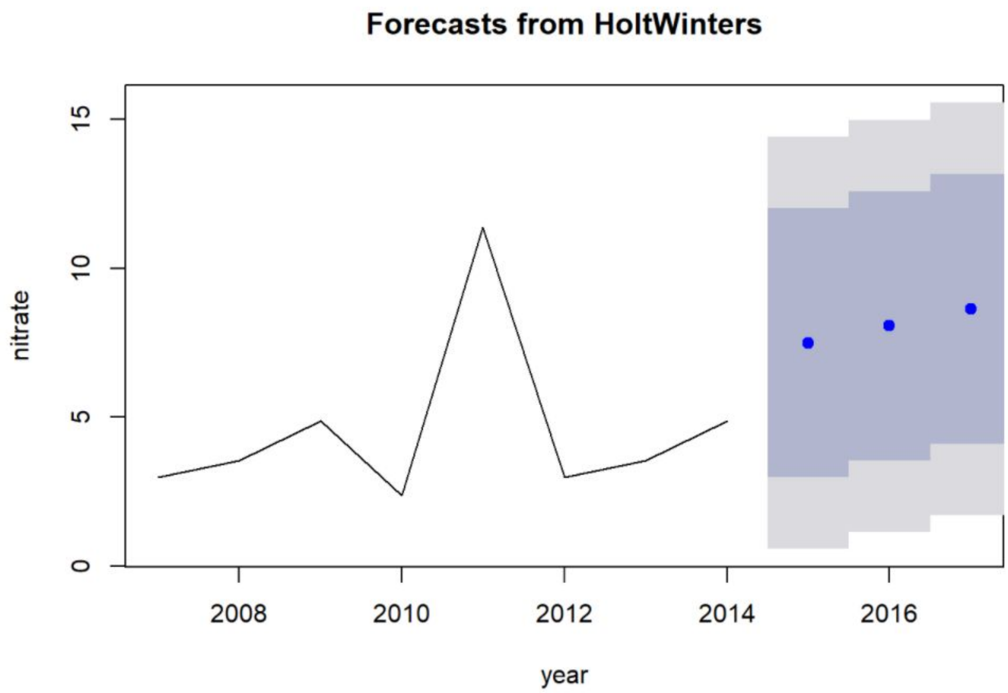
Maharashtra

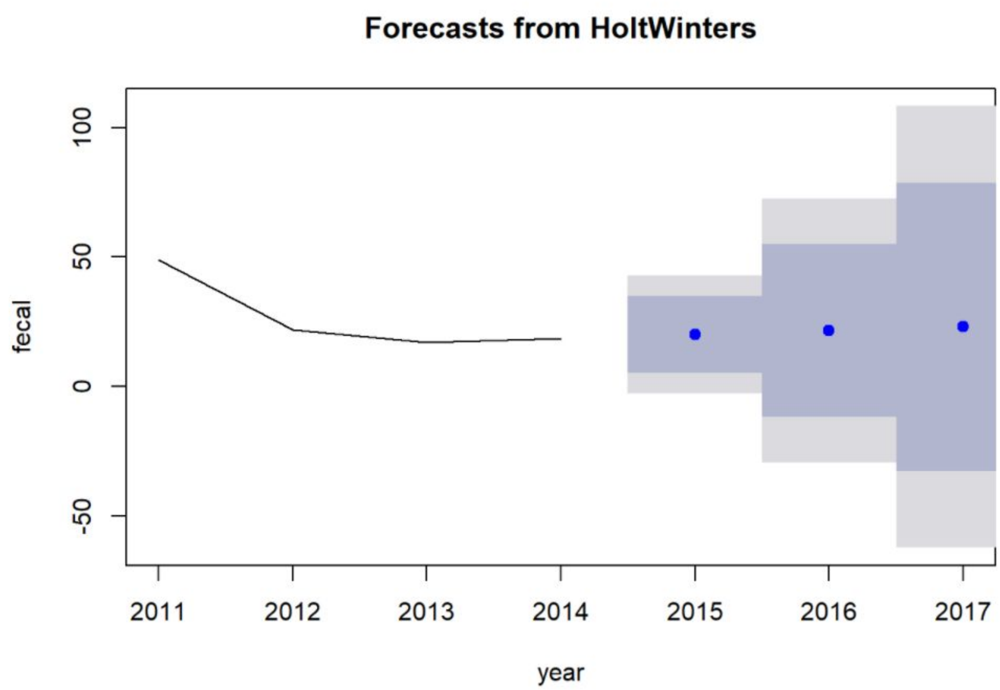





Forecasts from HoltWinters
conductivity
400
600
800
1000
2011
2012
2013
2014
2015
2016
2017
year

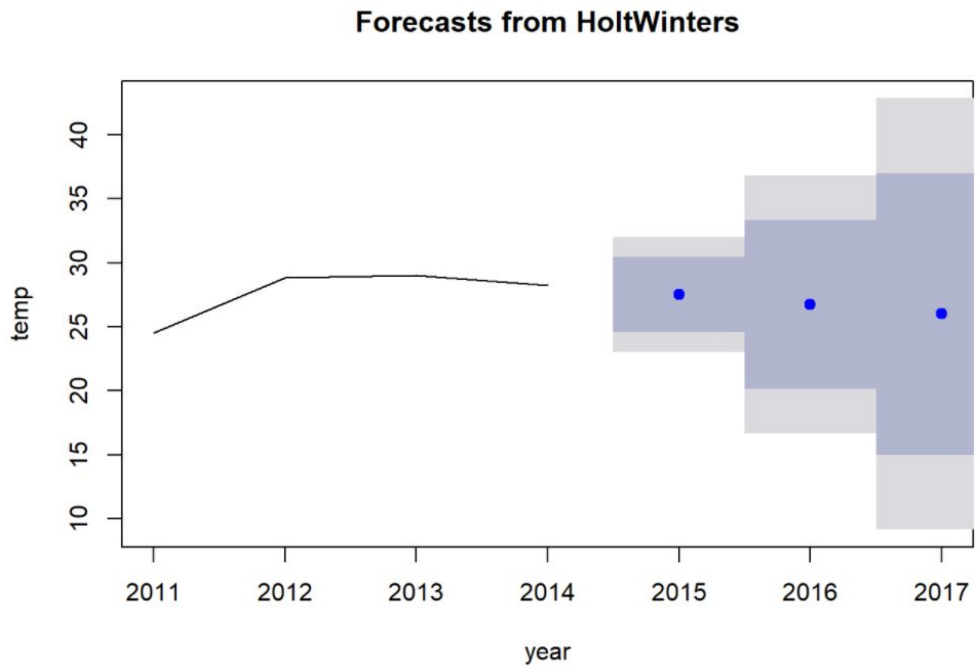

Forecasts from HoltWinters
BOD
-10
-5
0
5
10
15
2011
2012
2013
2014
2015
2016
2017
year

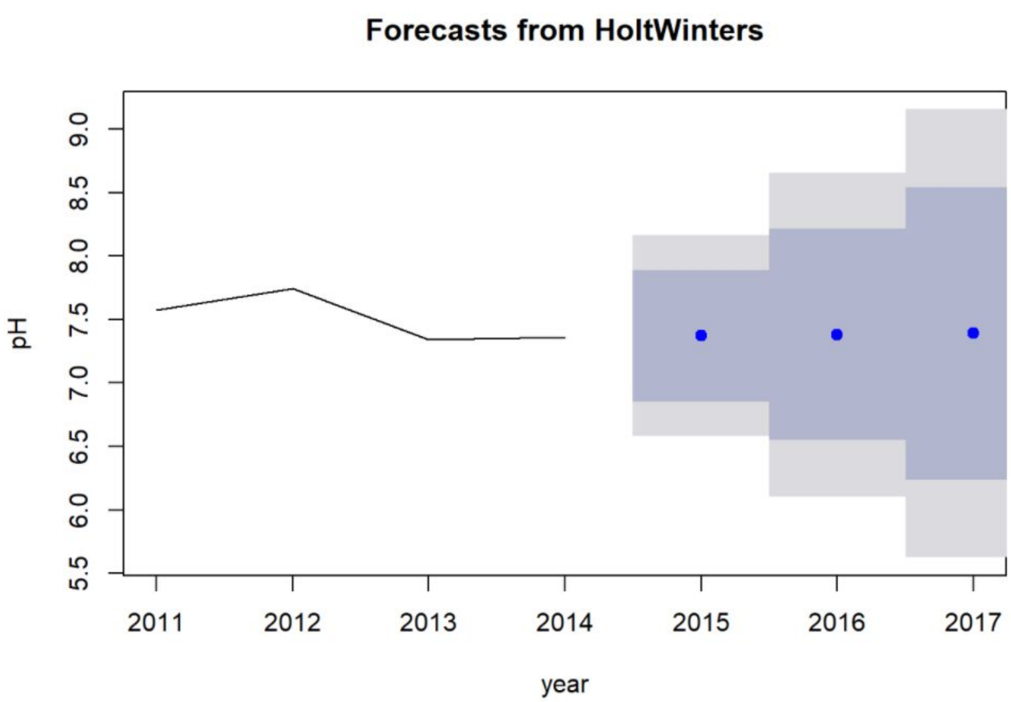

Forecasts from HoltWinters
nitrate
0
5
10
15
2008
2010
2012
2014
2016
year

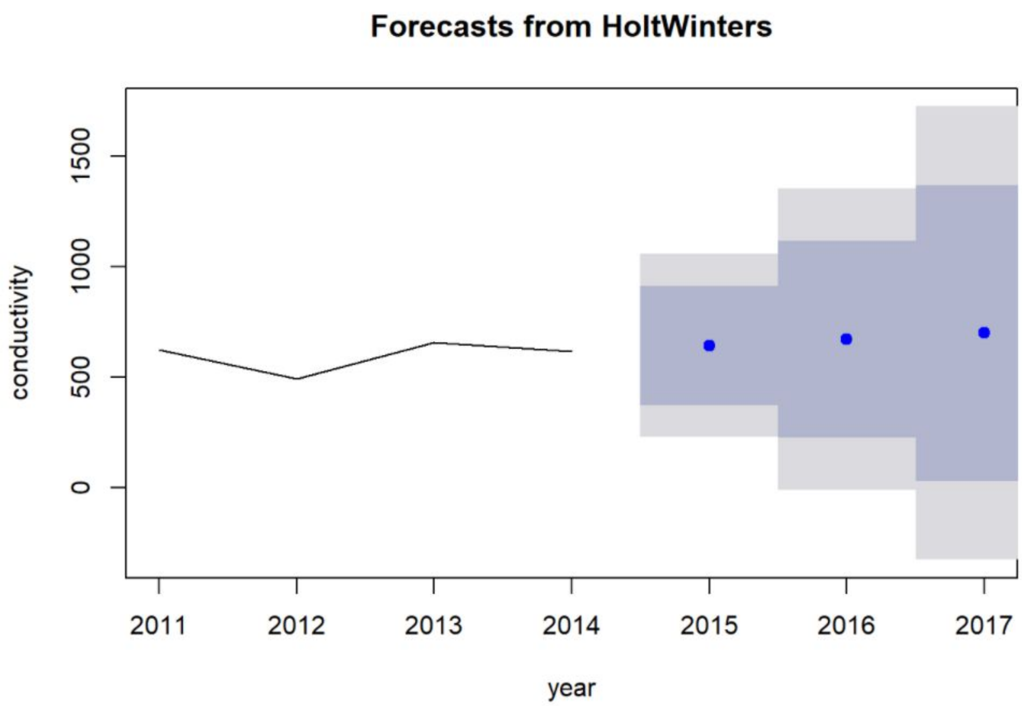

Forecasts from HoltWinters
fecal
-50
0
50
100
2011
2012
2013
2014
2015
2016
2017
year

Odisha

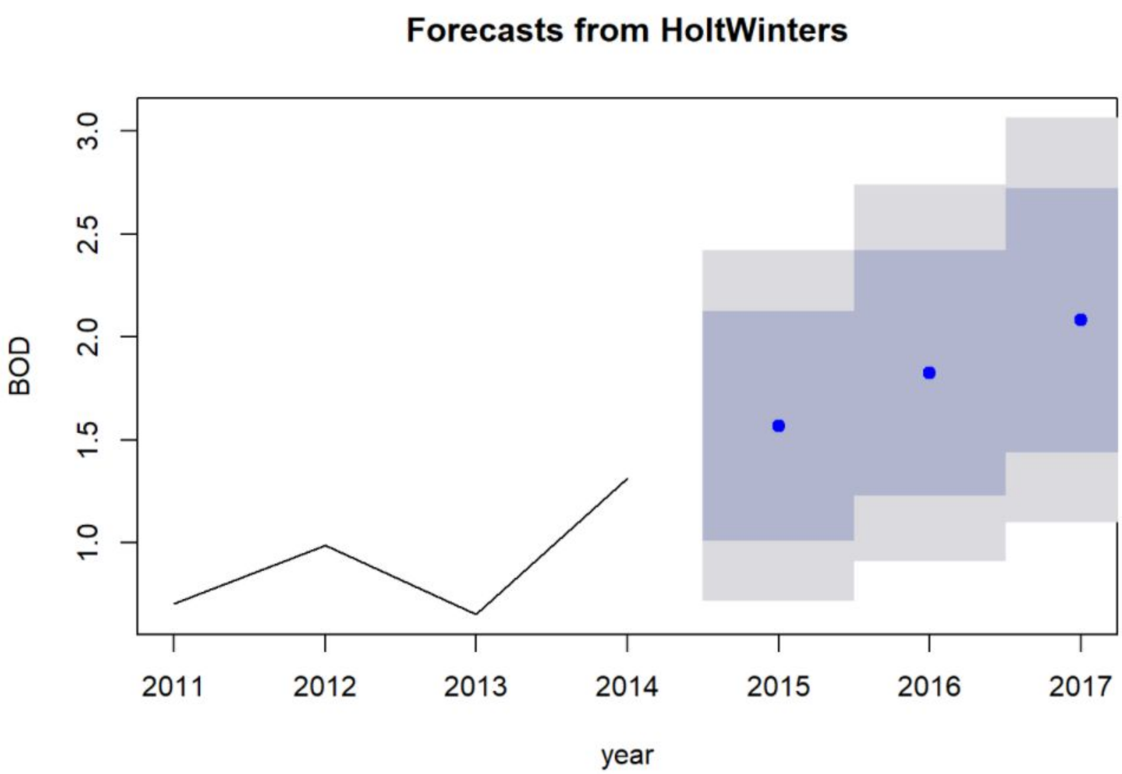
https://github.com/VDSouz/IndiaWaterQuality/blob/master/States/report_ODISHA_2018-01-14.html





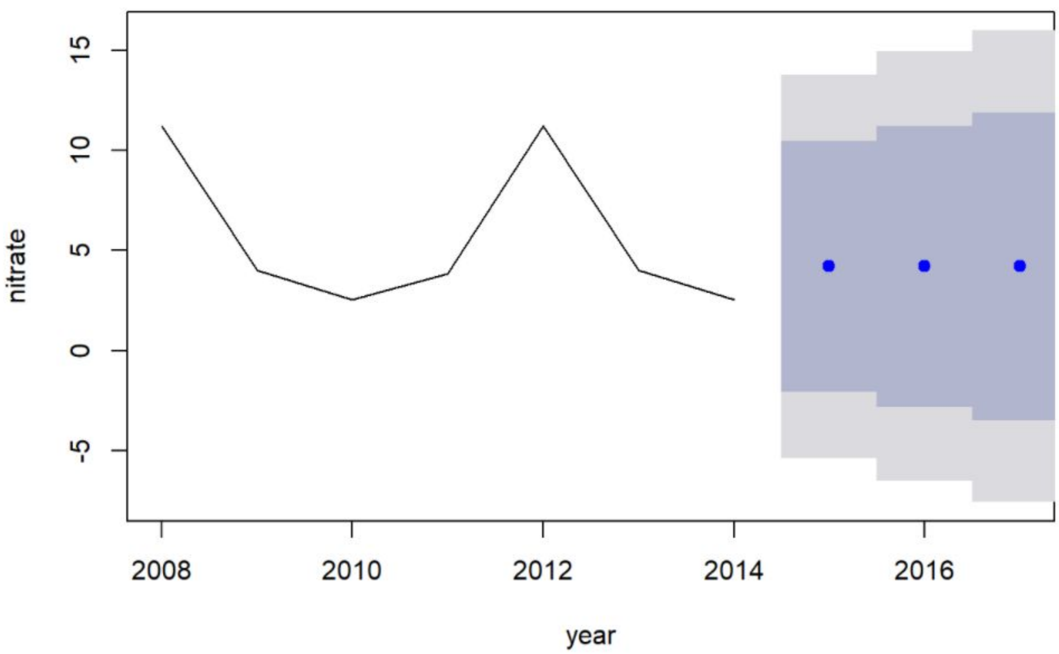


Forecasts from HoltWinters
conductivity
0
500
1000
1500
2011
2012
2013
2014
2015
2016
2017
year

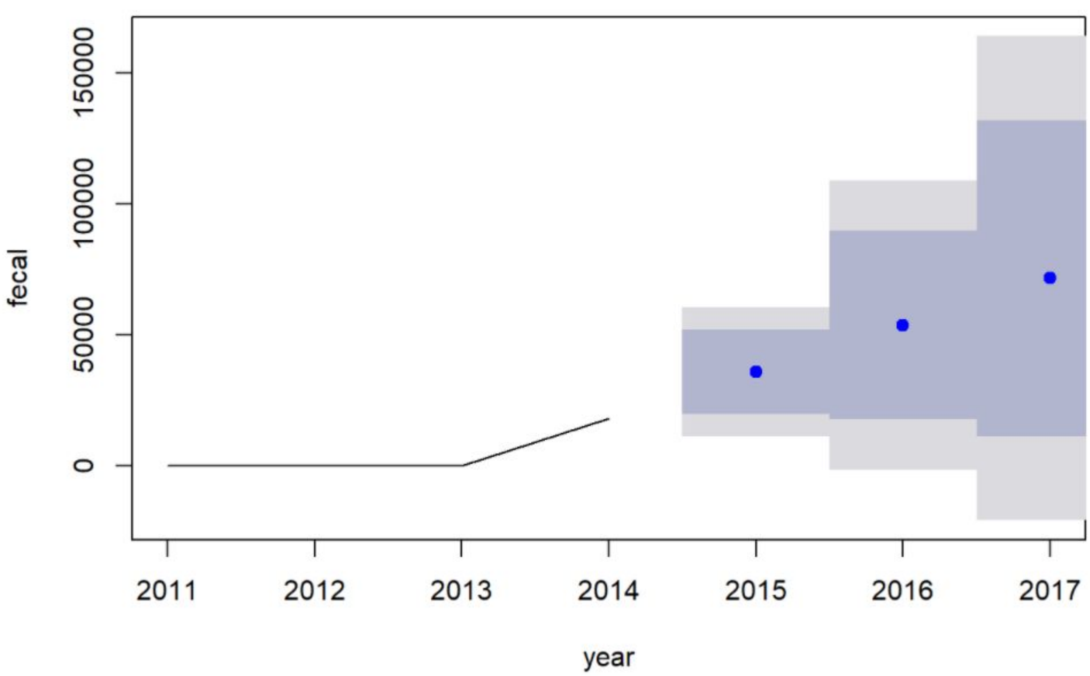


Forecasts from HoltWinters
BOD
1.0
1.5
2.0
2.5
3.0
2011
2012
2013
2014
2015
2016
2017
year

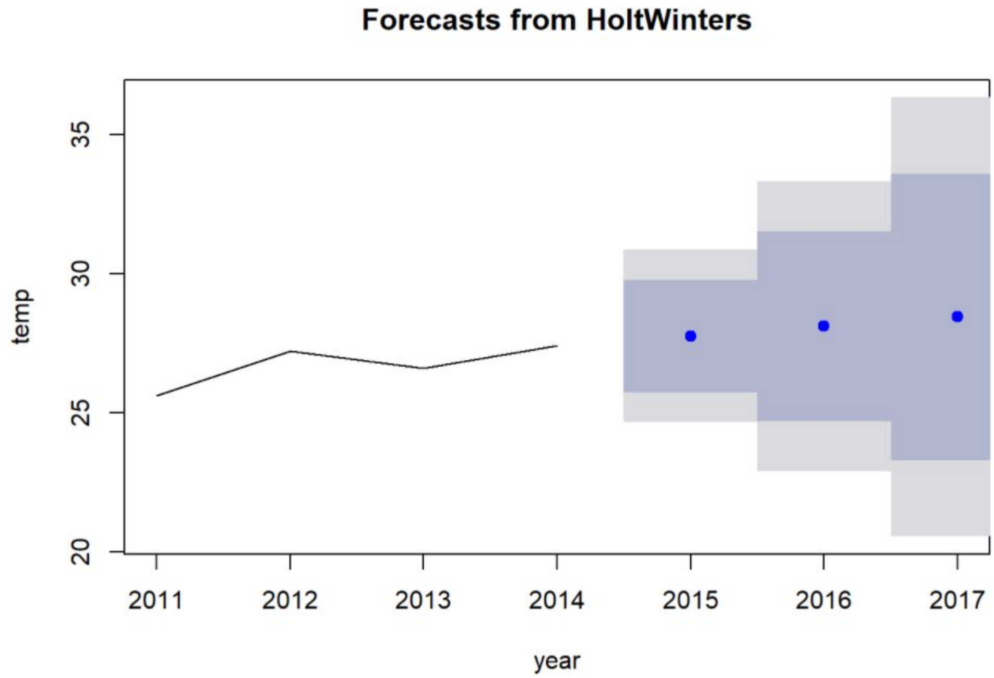
Forecasts from HoltWinters



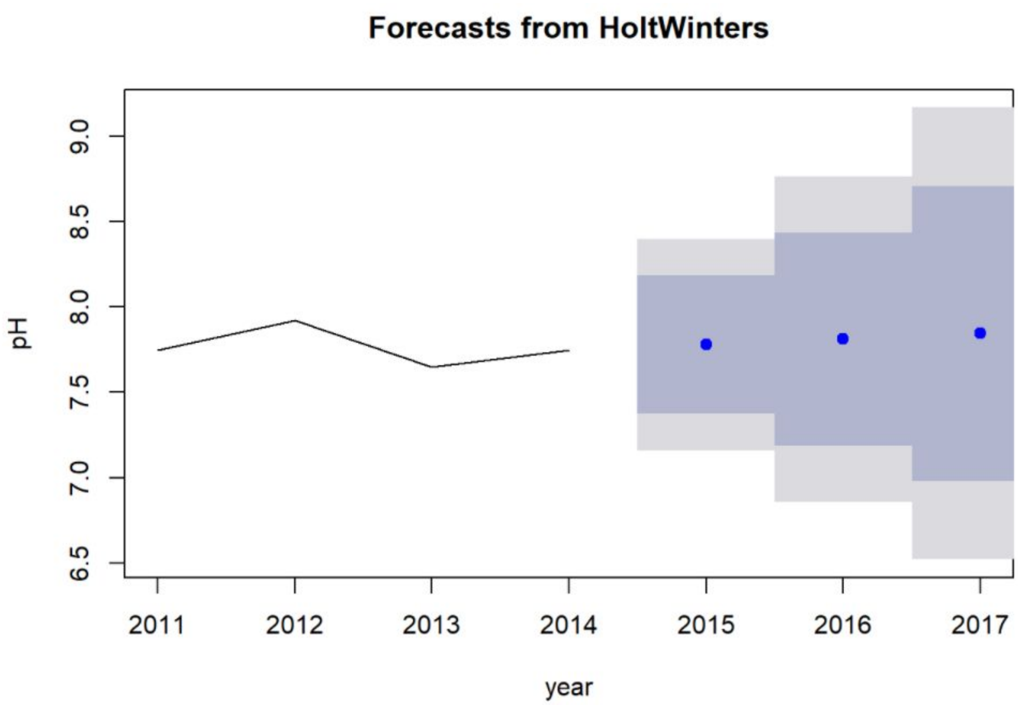
Forecasts from HoltWinters



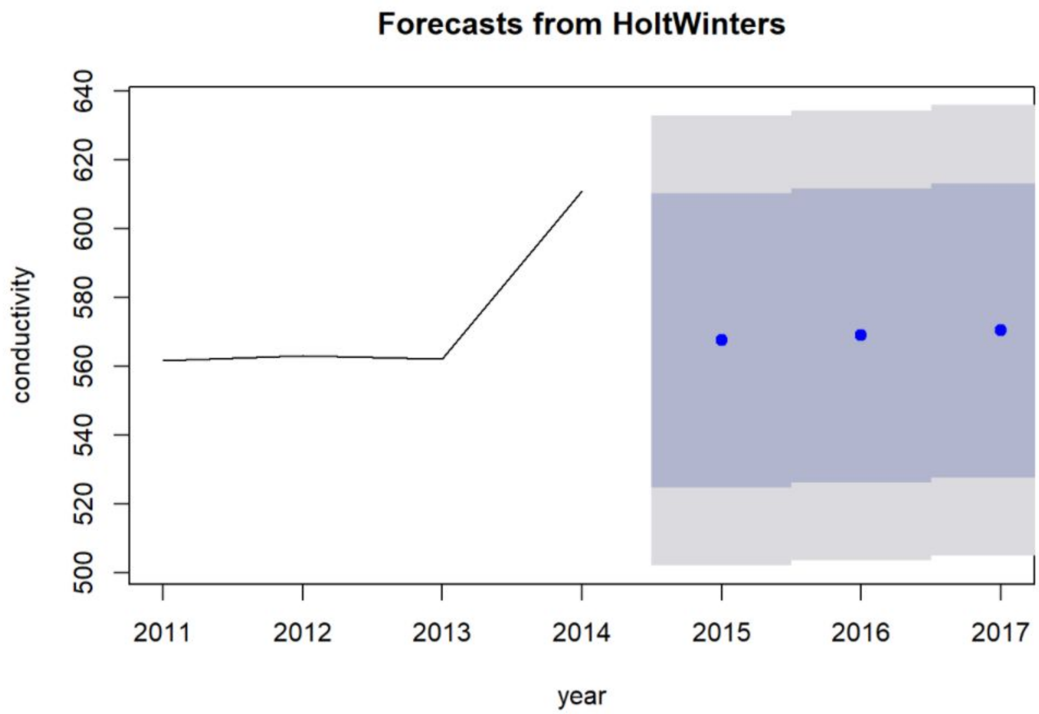
Rajasthan

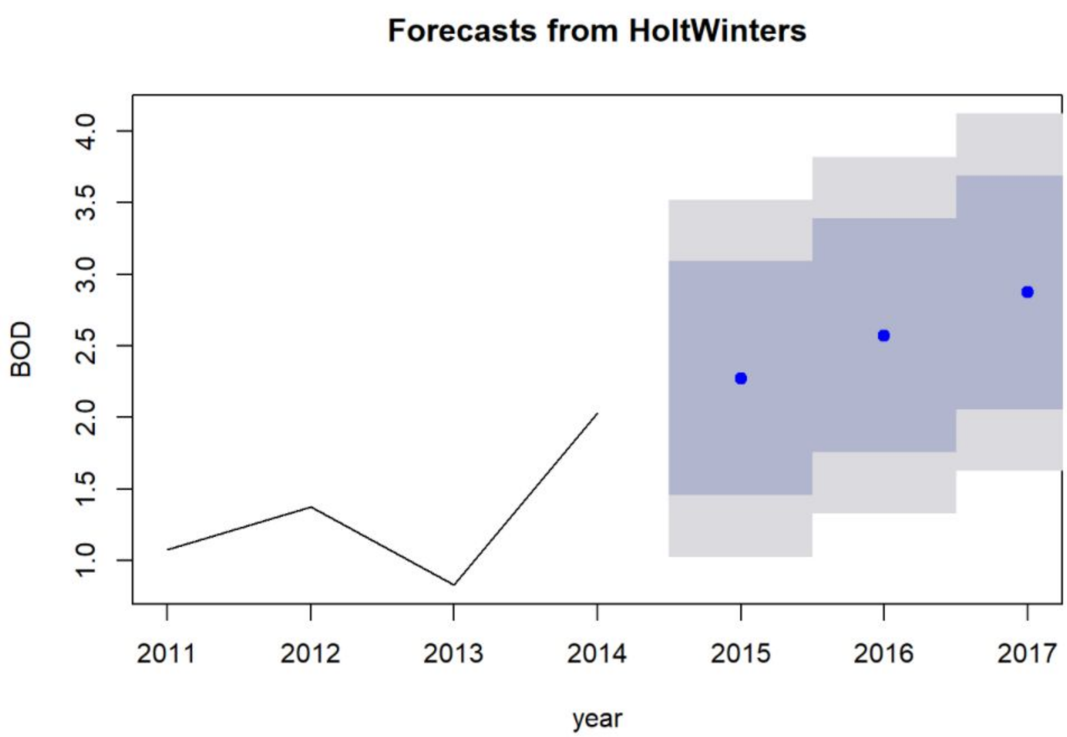


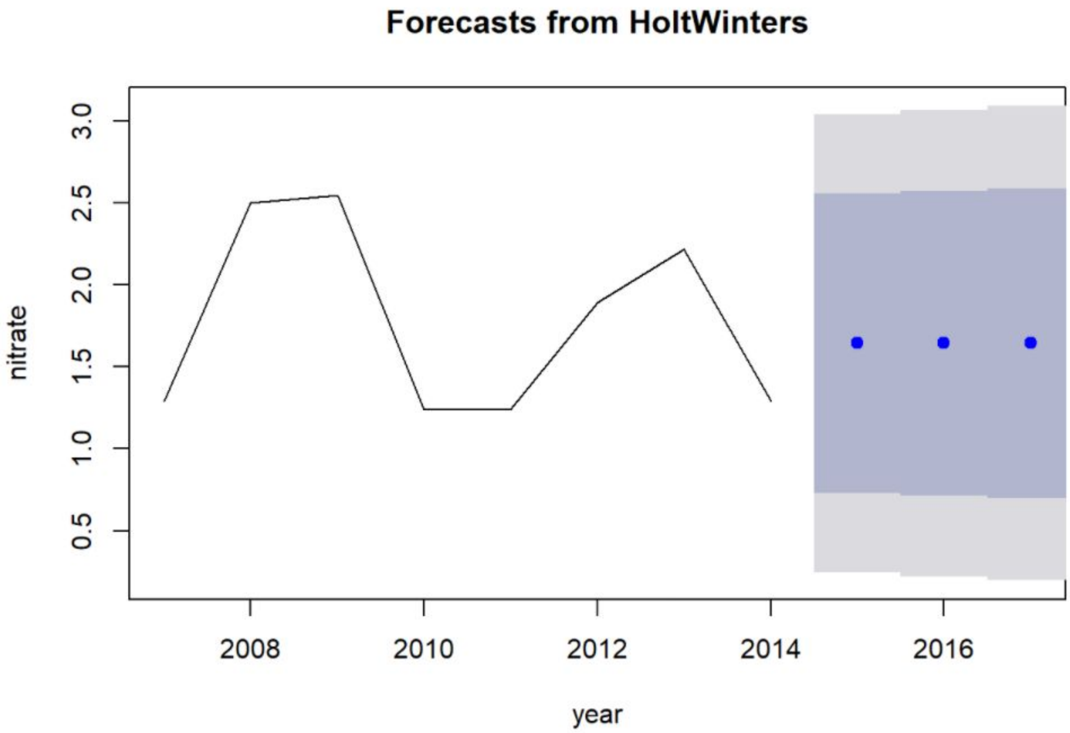
Forecasts from HoltWinters
temp
35
30
25
20
2011
2012
2013
2014
2015
2016
2017
year

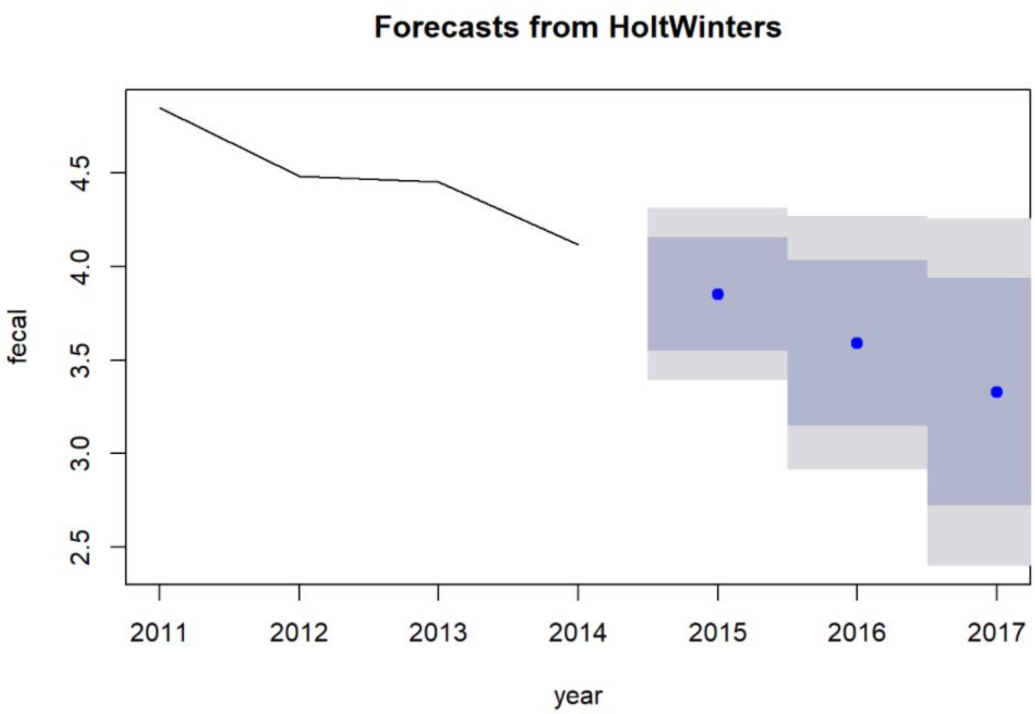


Forecasts from HoltWinters
pH
9.0
8.5
8.0
7.5
7.0
6.5
2011
2012
2013
2014
2015
2016
2017
year

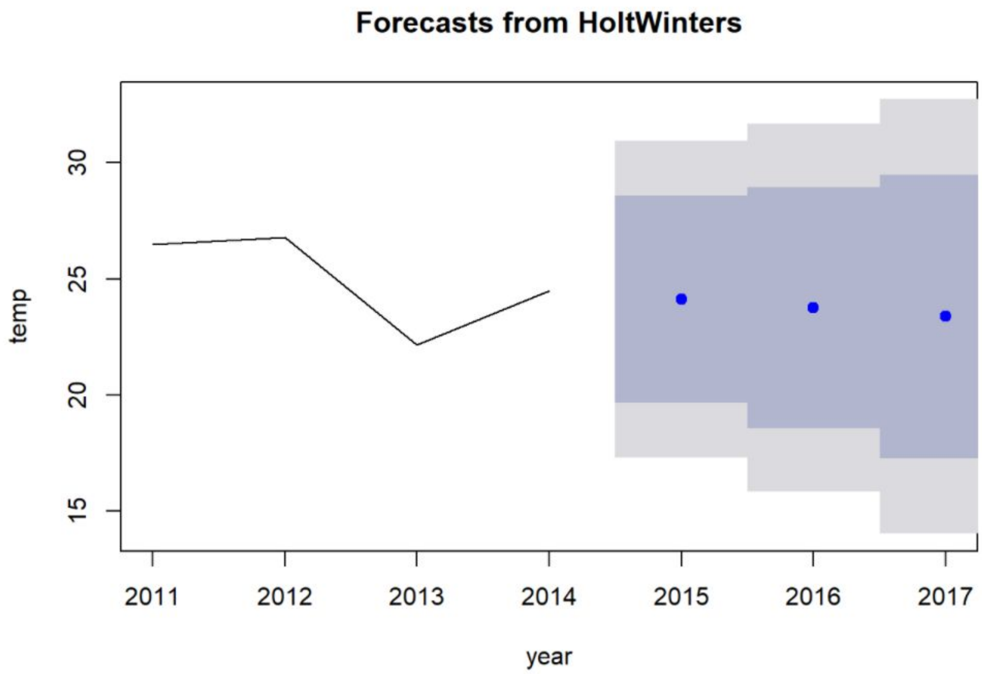

Forecasts from HoltWinters
conductivity
640
620
600
580
560
540
520
500
2011
2012
2013
2014
2015
2016
2017
year

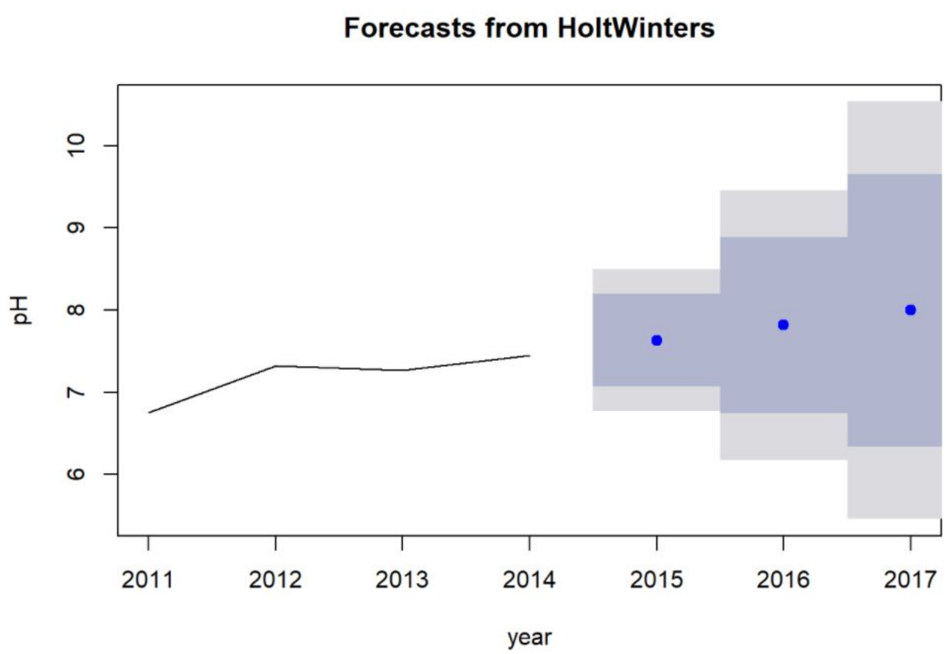

Forecasts from HoltWinters
BOD
4.0
3.5
3.0
2.5
2.0
1.5
1.0
2011
2012
2013
2014
2015
2016
2017
year


Forecasts from HoltWinters
nitrate
0.5
1.0
1.5
2.0
2.5
3.0
2008
2010
2012
2014
2016
year

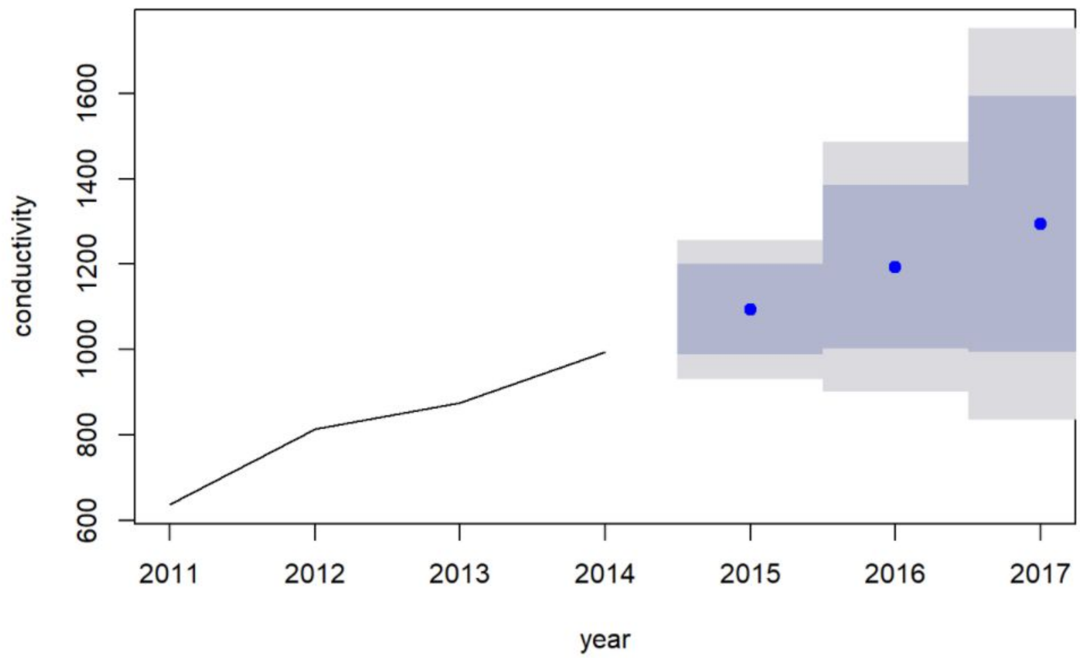

Forecasts from HoltWinters
fecal
2.5
3.0
3.5
4.0
4.5
2011
2012
2013
2014
2015
2016
2017
year

Uttar Pradesh

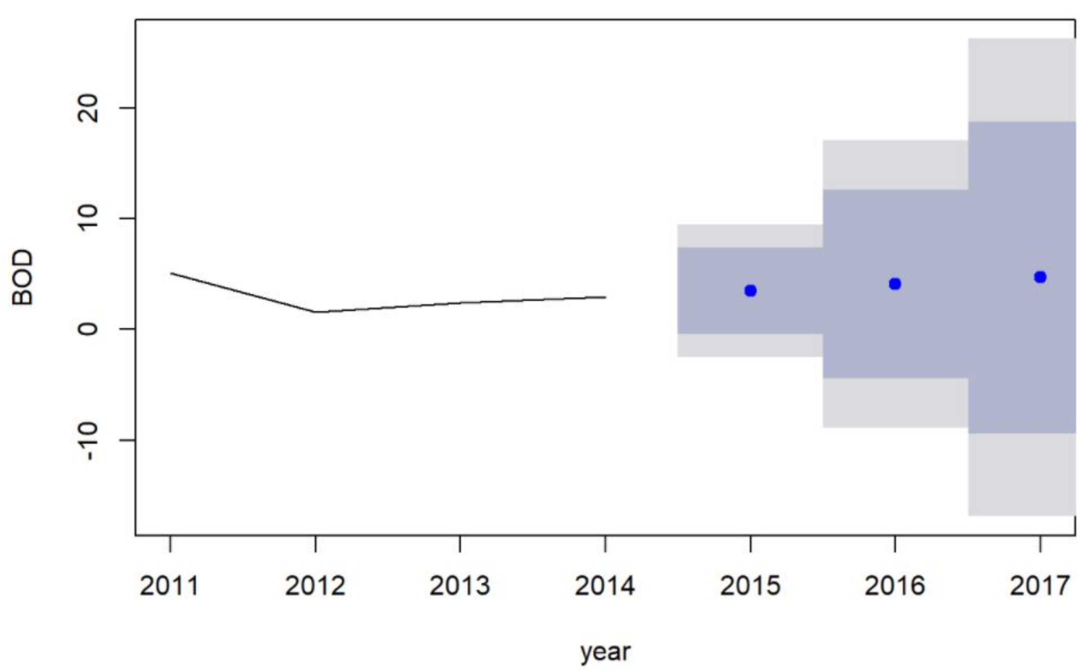




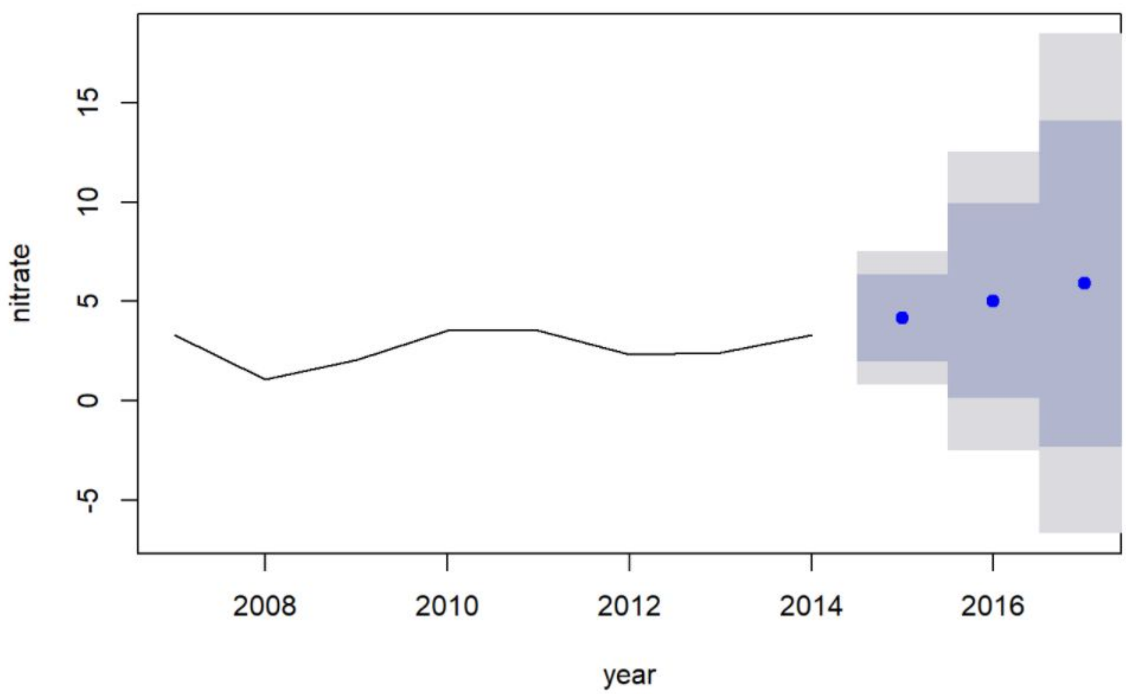
Forecasts from HoltWinters



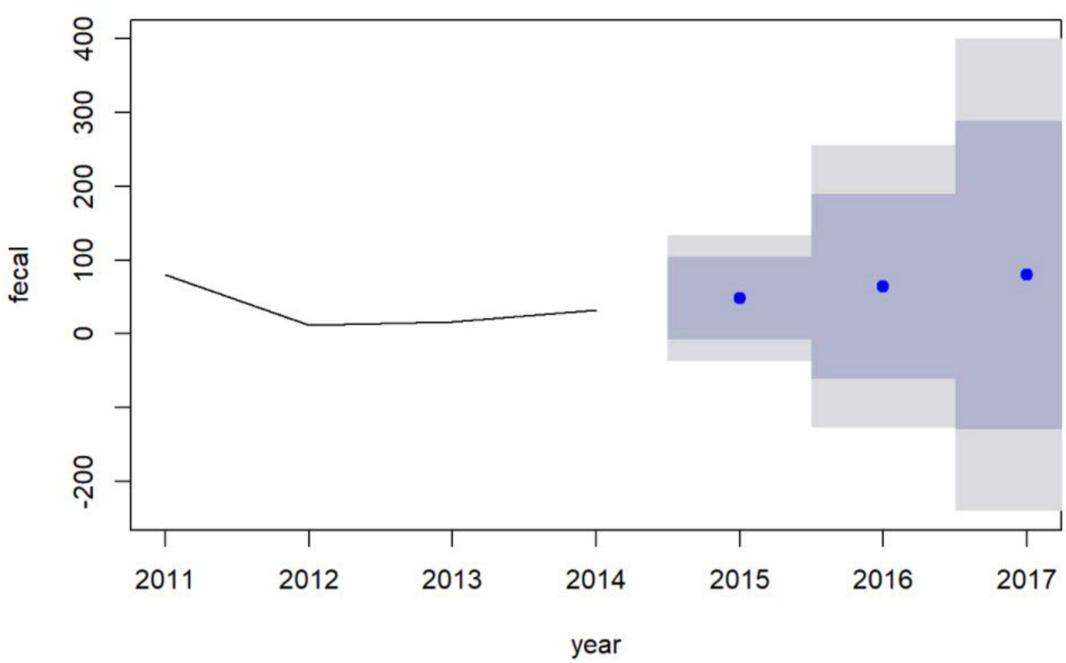
Forecasts from HoltWinters



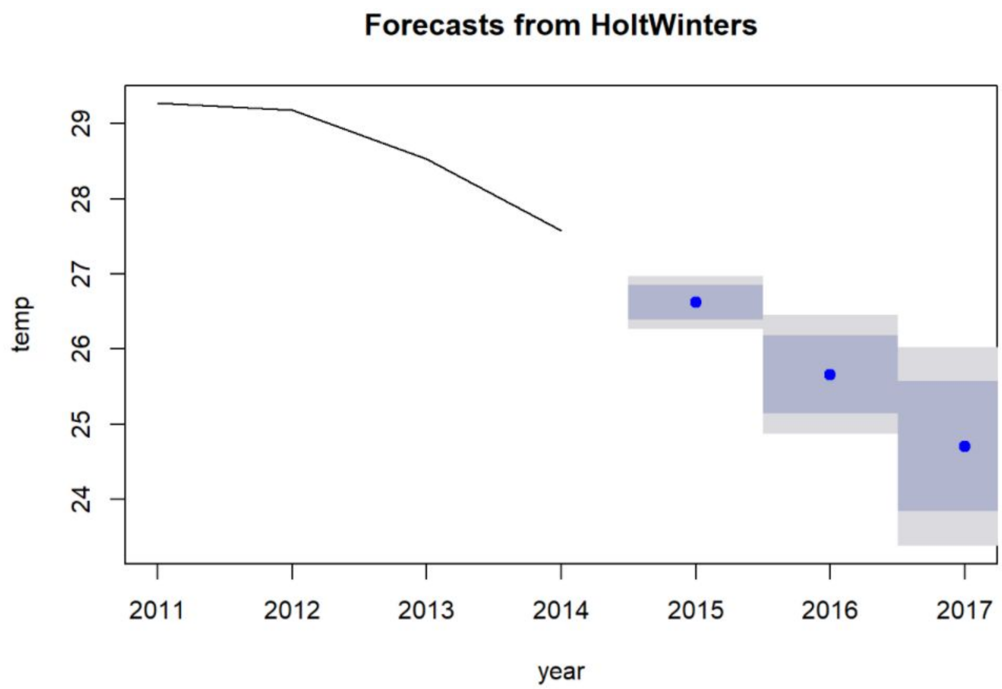
Forecasts from HoltWinters



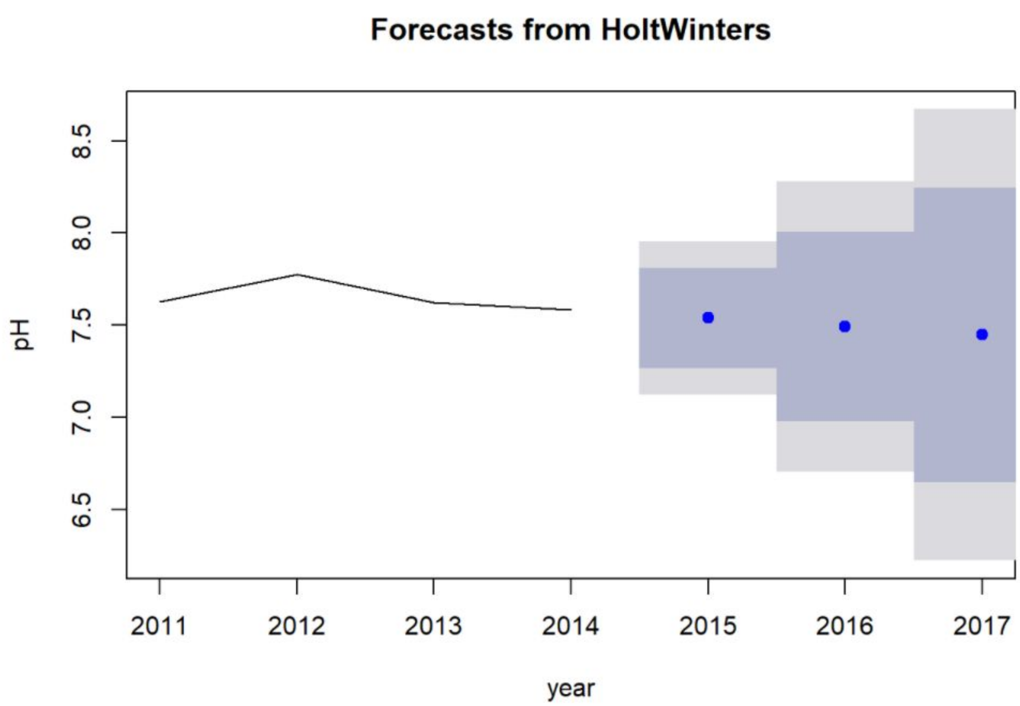
Forecasts from HoltWinters



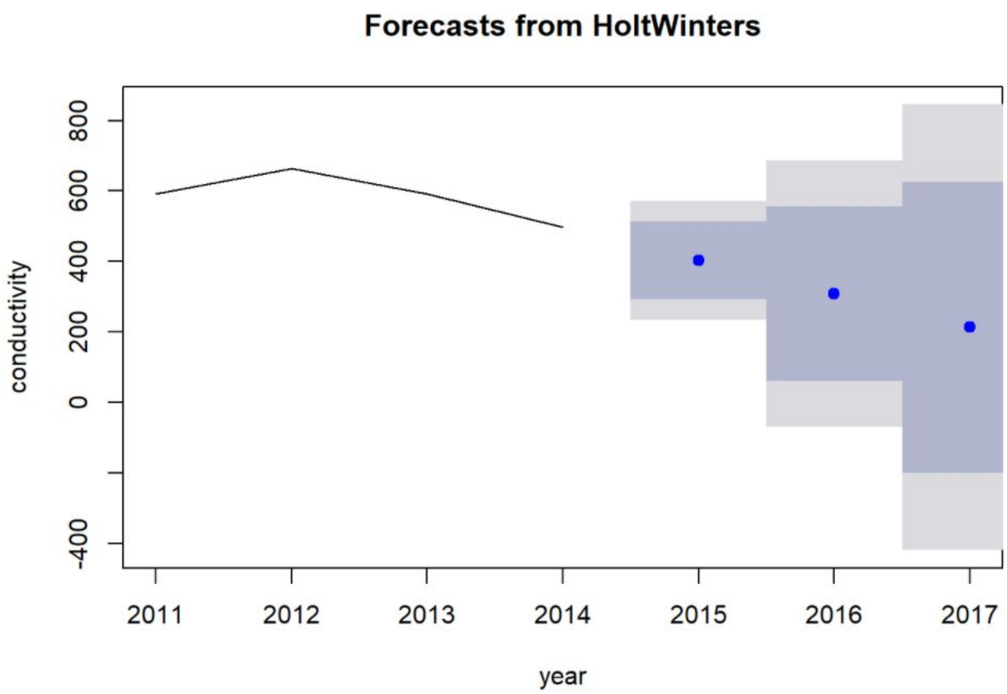
West Bengal

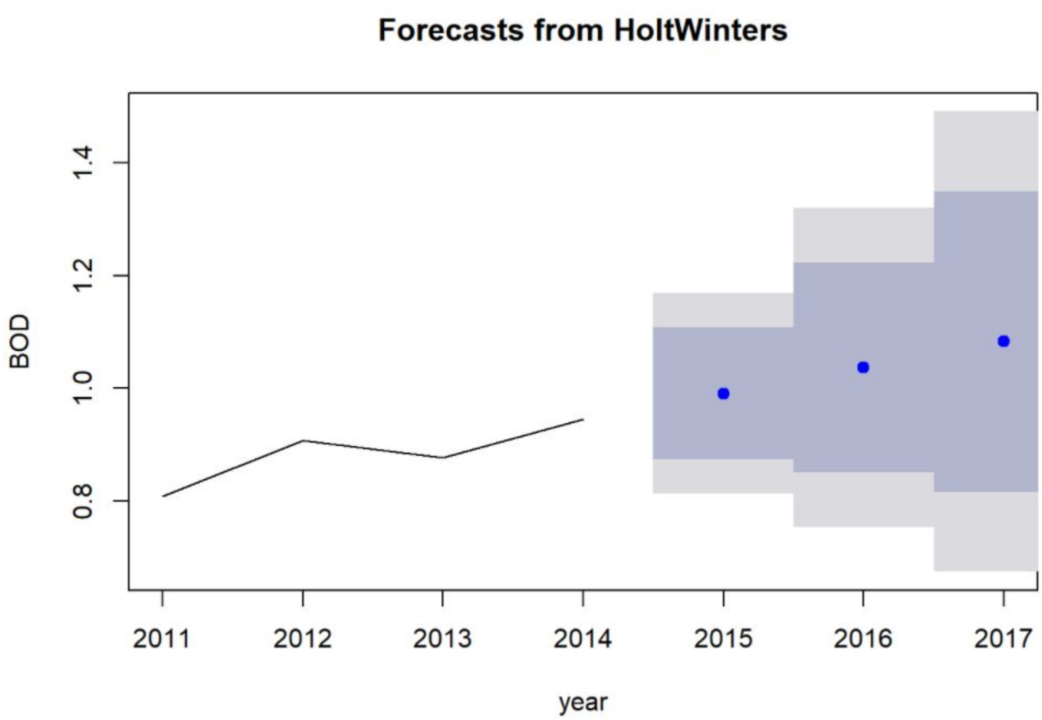


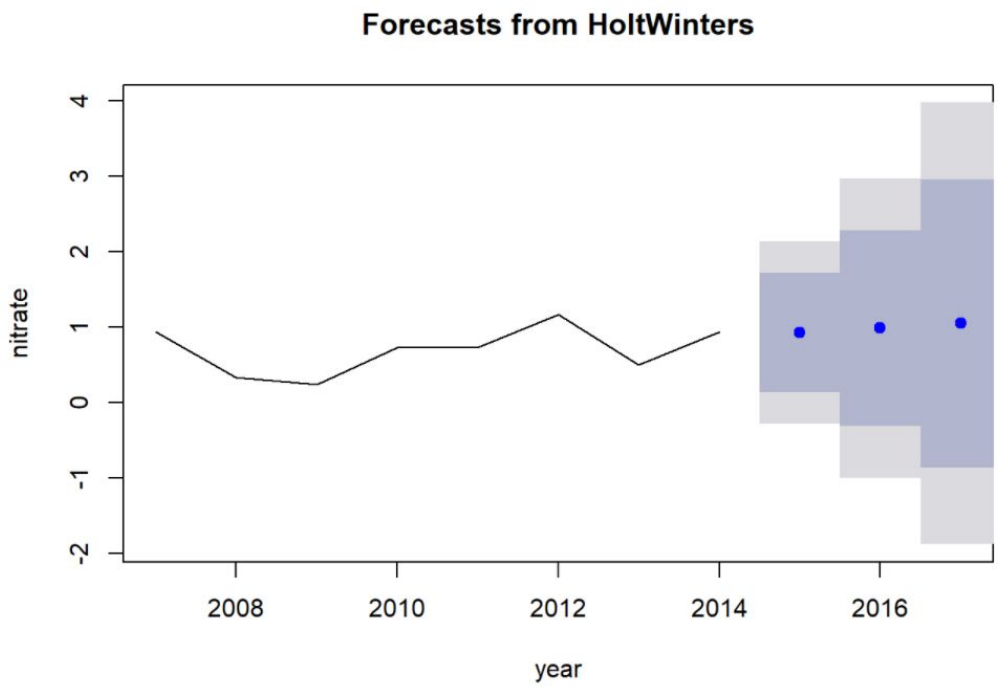
Forecasts from HoltWinters
temp
29
28
27
26
25
24
2011
2012
2013
2014
2015
2016
2017
year

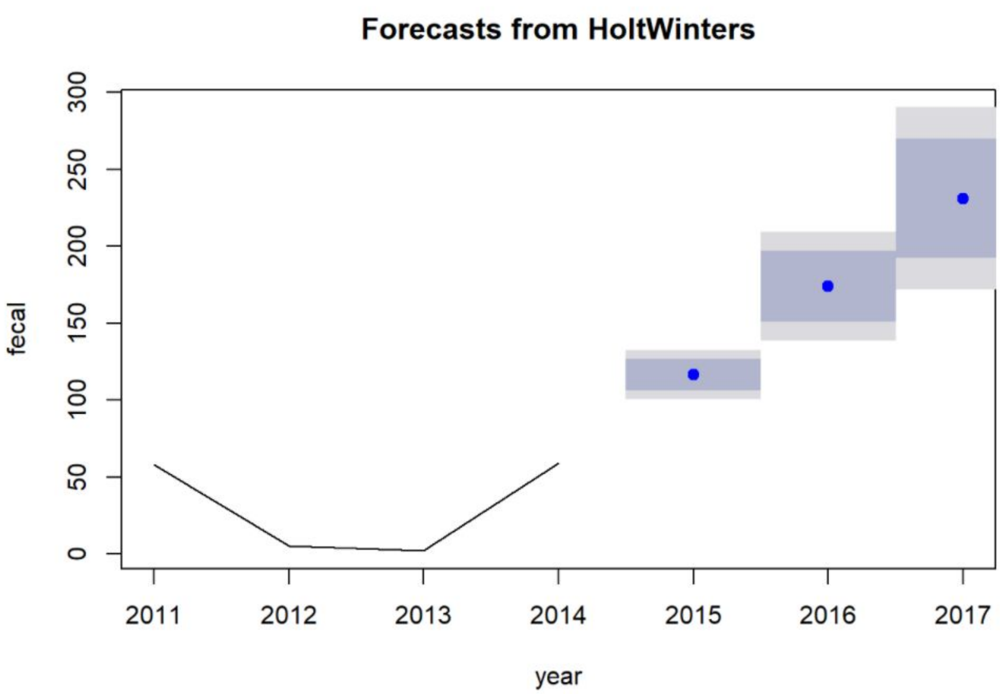


Forecasts from HoltWinters
pH
8.5
8.0
7.5
7.0
6.5
2011
2012
2013
2014
2015
2016
2017
year

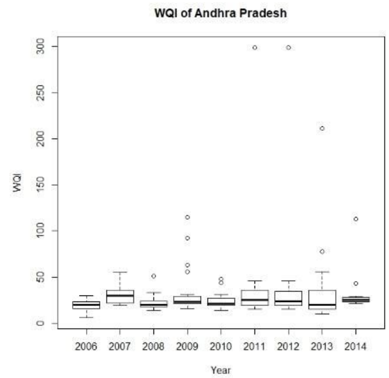

Forecasts from HoltWinters
conductivity
800
600
400
200
0
-400
2011
2012
2013
2014
2015
2016
2017
year

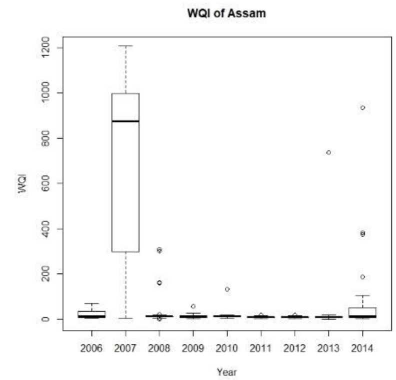

Forecasts from HoltWinters
BOD
1.4
1.2
1.0
0.8
2011
2012
2013
2014
2015
2016
2017
year

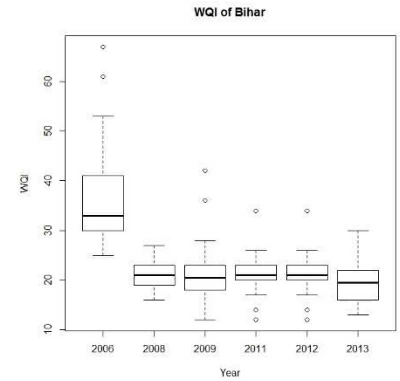

Forecasts from HoltWinters
nitrate
4
3
2
1
0
-1
-2
2008
2010
2012
2014
2016
year

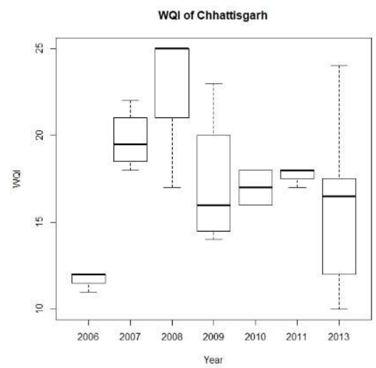

Forecasts from HoltWinters
fecal
300
250
200
150
100
50
0
2011
2012
2013
2014
2015
2016
2017
year

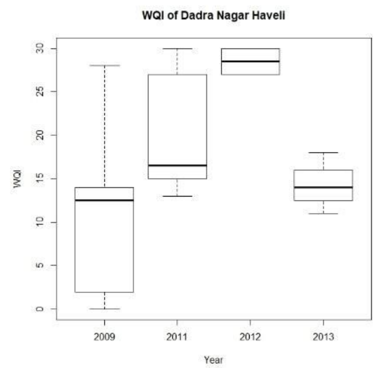
BoxPlots

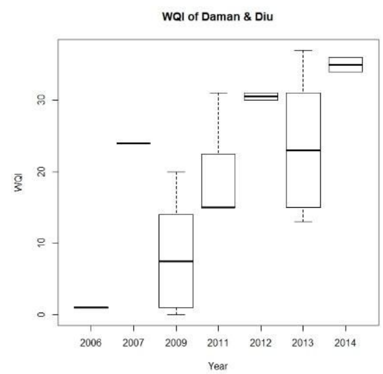

WQI of Andhra Pradesh
WQI
Year
2006 2007 2008 2009 2010 2011 2012 2013 2014

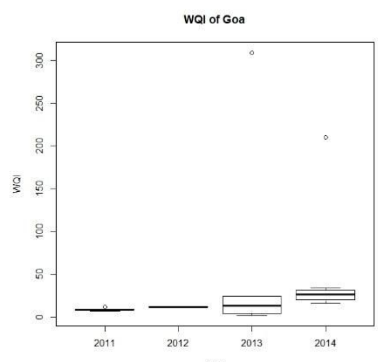

WQI of Assam
WQI
Year
2006 2007 2008 2009 2010 2011 2012 2013 2014

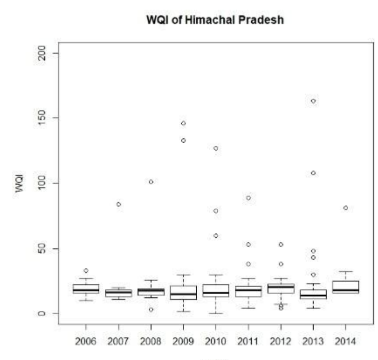

WQI of Bihar
WQI
Year
2006 2008 2009 2011 2012 2013

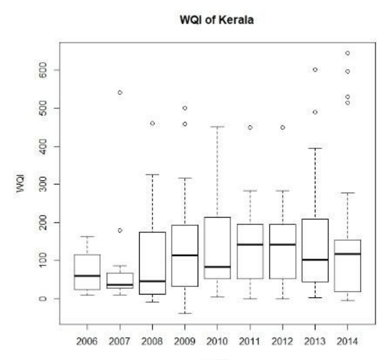

WQI of Chhattisgarh
WQI
Year
2006 2007 2008 2009 2010 2011 2013

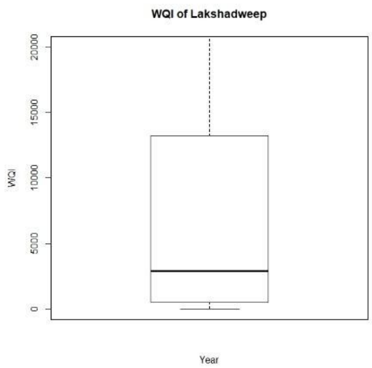

WQI of Dadra Nagar Haveli
WQI
Year
2009 2011 2012 2013

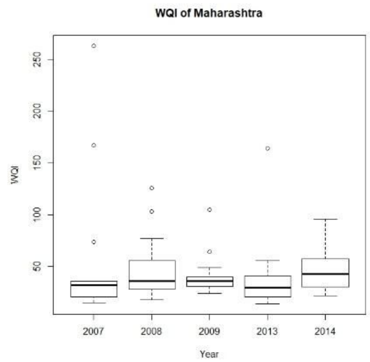

WQI of Daman & Diu
WQI
Year
2006 2007 2009 2011 2012 2013 2014

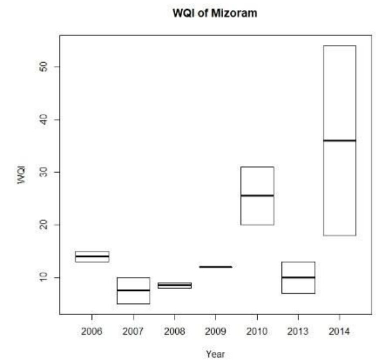

WQI of Goa
WQI
Year
2011 2012 2013 2014

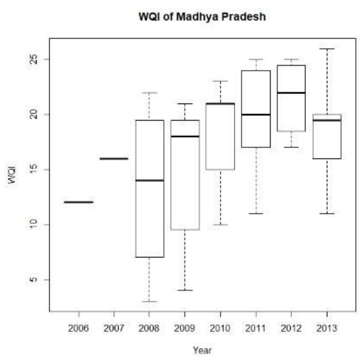

WQI of Himachal Pradesh
WQI
Year
2006 2007 2008 2009 2010 2011 2012 2013 2014

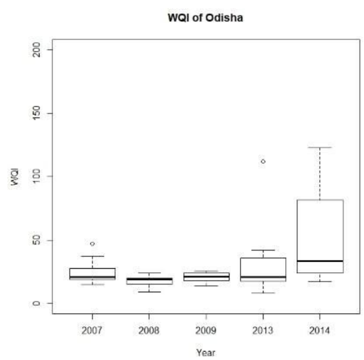

WQI of Kerala
WQI
Year
2006 2007 2008 2009 2010 2011 2012 2013 2014

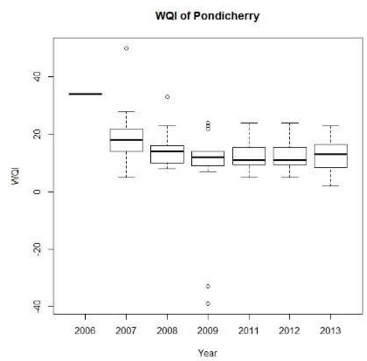

WQI of Lakshadweep
WQI
Year

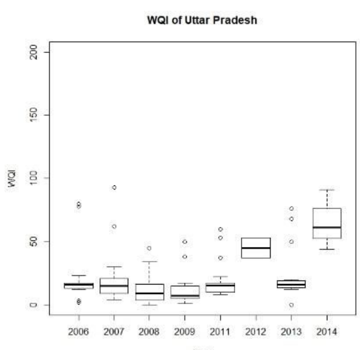

WQI of Maharashtra
WQI
2007 2008 2009 2013 2014
Year

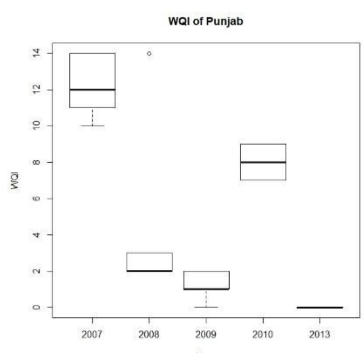

WQI of Mizoram
WQI
2006 2007 2008 2009 2010 2013 2014
Year

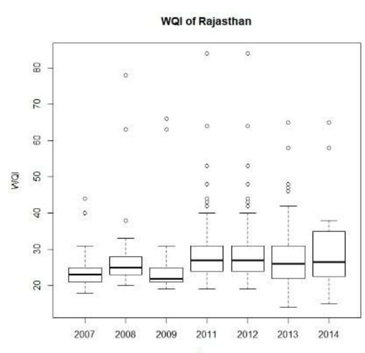

WQI of Madhya Pradesh
WQI
2006 2007 2008 2009 2010 2011 2012 2013
Year

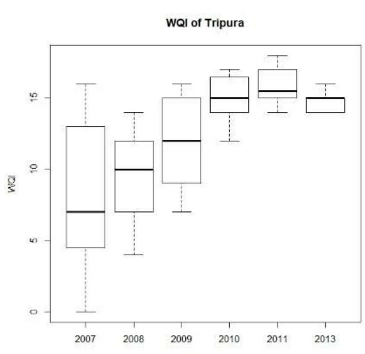

WQI of Odisha
WQI
2007 2008 2009 2013 2014
Year

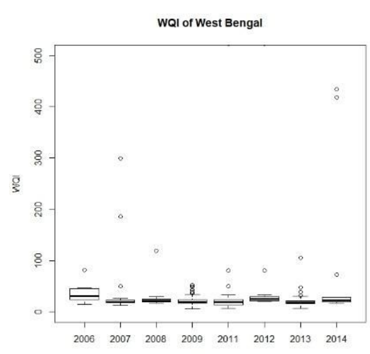

WQI of Pondicherry
WQI
2006 2007 2008 2009 2011 2012 2013
Year


WQI of Uttar Pradesh
WQI
2006 2007 2008 2009 2011 2012 2013 2014
Year


WQI of Punjab
WQI
2007 2008 2009 2010 2013
Year


WQI of Rajasthan
WQI
2007 2008 2009 2011 2012 2013 2014
Year


WQI of Tripura
WQI
2007 2008 2009 2010 2011 2013
Year


WQI of West Bengal
WQI
2006 2007 2008 2009 2011 2012 2013 2014
Year

R code

https://github.com/VDSouz/IndiaWaterQuality

config.R

```
# File 1:R-Script
# contains a loop that iteratively calls an Rmarkdown file (i.e. File 2)

# load packages
library(knitr)
library(markdown)
library(rmarkdown)
library(readxl)

#Load Water Quality Data File
data <- read.csv("C:/Users/vanes/Documents/Report/waterqualitydata.csv")

# remove NA
row.has.na <- apply(data, 1, function(x){any(is.na(x))})
sum(row.has.na)
final_data <- data[!row.has.na,]
head(final_data)
data <- final_data

# for each type of state in the data - created a report
# these reports are saved in output_dir with the name specified by
output_file
for (state in unique(data$STATE)){

  if (state == "ANDHRA PRADESH" ||    state ==  "ASSAM" || state == "GOA"
|| state == "HIMACHAL PRADESH" || state == "KERALA" || state ==
"MAHARASHTRA" || state == "ODISHA" || state == "RAJASTHAN" || state ==
"UTTAR PRADESH" || state == "WEST BENGAL")
  rmarkdown::render('C:/Users/vanes/Documents/Report/State-wise.Rmd',  #
file 2
 output_file =  paste("report_", state, '_', Sys.Date(), ".html", sep=''),
 output_dir = 'C:/Users/vanes/Documents/Report/States')
}
```

state-wise.Rmd

```
---
title: "State wise Analysis"
output: html_document
---

## The Data
The data is extracted from https://data.gov.in/sector/water-quality
```{r}
# run report for current car listed in loop in R file
s1 <- data[data$STATE==state,]
datatable(s1)
```

Parameters - Mean values
Water quality data is collected across different areas within a state.
Therefore calculate mean values for the 6 parameters
```{r}
temp <-
ddply(s1,~YEAR,summarise,mean=mean(as.numeric(as.double(mean_temp))))
ph <- ddply(s1,~YEAR,summarise,mean=mean(as.numeric(as.double(mean_ph))))
conductivity <-
ddply(s1,~YEAR,summarise,mean=mean(as.numeric(as.double(mean_conductivity))
))
bod <- ddply(s1,~YEAR,summarise,mean=mean(as.numeric(as.double(mean_bod))))
nitrate <-
ddply(s1,~YEAR,summarise,mean=mean(as.numeric(as.double(mean_nitratenitrite
))))
fecal <-
ddply(s1,~YEAR,summarise,mean=mean(as.numeric(as.double(mean_fecalcoliform)
)))
```

Parameter
The 6 parameter values are determined at a state level for the specified
time range
```{r}
temp1 <- temp$mean
ph1 <- ph$mean
conductivity1 <- conductivity$mean
```
```

````
bod1 <- bod$mean
nitrate1 <- nitrate$mean
fecal1 <- fecal$mean
```

## TimeSeries
Time series data may be defined as an ordered sequence of values of a
variable at equally spaced time intervals.
Based on Data Available we have an dynamic or manual time range..
Time series is created for each paramter
```{r}
if(state == 'MAHARASHTRA' | state == 'ODISHA' | state == 'RAJASTHAN' |
state == 'UTTAR PRADESH' | state == 'WEST BENGAL'){
## custom
    temp_ts <-ts(temp1, start=c(2011, 1), end=c(2014, 1), frequency=1)
    temp_ts
    ph_ts <-ts(ph1, start=c(2011, 1), end=c(2014, 1), frequency=1)
    ph_ts
    conductivity_ts <-ts(conductivity1, start=c(2011, 1), end=c(2014, 1),
frequency=1)
    conductivity_ts
    bod_ts <- ts(bod1, start=c(2011, 1), end=c(2014, 1), frequency=1)
    bod_ts
    nitrate_ts <- ts(nitrate1,  start=c(min(nitrate$YEAR), 1),
end=c(max(nitrate$YEAR),1), frequency=1)
    nitrate_ts
    fecal_ts <- ts(fecal1,  start=c(2011, 1), end=c(2014, 1), frequency=1)
    fecal_ts
} else {
##Dynamic
  temp_ts <-ts(temp1,  start=c(min(temp$YEAR), 1), end=c(max(temp$YEAR),1),
frequency=1)
  temp_ts
  ph_ts <-ts(ph1,  start=c(min(ph$YEAR), 1), end=c(max(ph$YEAR),1),
frequency=1)
  ph_ts
  conductivity_ts <-ts(conductivity1, start=c(min(conductivity$YEAR), 1),
end=c(max(conductivity$YEAR),1), frequency=1)
  conductivity_ts
  bod_ts <- ts(bod1,  start=c(min(bod$YEAR), 1), end=c(max(bod$YEAR),1),
frequency=1)
````

````
  bod_ts
  nitrate_ts <- ts(nitrate1,  start=c(min(nitrate$YEAR), 1),
end=c(max(nitrate$YEAR),1), frequency=1)
  nitrate_ts
  fecal_ts <- ts(fecal1,  start=c(min(fecal$YEAR), 1),
end=c(max(fecal$YEAR),1), frequency=1)
  fecal_ts
}
```



## TimeSeries Plots
The time series plots are analysed
```{r}
plot(temp_ts)
plot(ph_ts)
plot(conductivity_ts)
plot(bod_ts)
plot(nitrate_ts)
plot(fecal_ts)
```



# Holt-Winters Exponential Smoothing
Holt-Winters exponential smoothing used to make short-term forecasts. It
estimates the level, slope and seasonal component at the current time
point.

The 6 paramters are forecasted and analysed using Holt-Winters exponential
smoothing

## Temprature
Temperature of the groundwater
```{r}
fit_temp_ts <- HoltWinters(temp_ts, gamma=FALSE)
fit_temp_ts
pred_temp_ts <- forecast(fit_temp_ts,3)

flag <- FALSE
for ( i in pred_temp_ts$mean)  if ( i < 0) flag <- TRUE
if (flag == TRUE)
````

```
  fit_temp_ts <- HoltWinters(temp_ts, beta = FALSE, gamma=FALSE)
  fit_temp_ts
  pred_temp_ts <- forecast(fit_temp_ts,3)

plot(pred_temp_ts)
pred_temp_ts

```

```
## pH
pH level of the groundwater (acceptable level: 6.5-8.5)
```{r}
fit_ph_ts <- HoltWinters(ph_ts, gamma=FALSE)
fit_ph_ts
pred_ph_ts <- forecast(fit_ph_ts,3)


flag <- FALSE
for ( i in pred_ph_ts$mean)  if ( i < 0) flag <- TRUE
if (flag == TRUE)
  fit_ph_ts <- HoltWinters(ph_ts, beta = FALSE, gamma=FALSE)
  fit_ph_ts
  pred_ph_ts <- forecast(fit_ph_ts,3)

plot(pred_ph_ts)
pred_ph_ts

```

Conductivity
Ability of groundwater to conduct current
```{r}
fit_conductivity_ts <- HoltWinters(conductivity_ts, gamma=FALSE)
fit_conductivity_ts
pred_conductivity_ts <- forecast(fit_conductivity_ts,3)


flag <- FALSE
for ( i in pred_conductivity_ts$mean)  if ( i < 0) flag <- TRUE
if (flag == TRUE)
  fit_conductivity_ts <- HoltWinters(conductivity_ts,beta = FALSE,
```
```

````
gamma=FALSE)
  fit_conductivity_ts
  pred_conductivity_ts <- forecast(fit_conductivity_ts,3)

plot(pred_conductivity_ts)
pred_conductivity_ts
```

## BOD
Amount of dissolved oxygen in the groundwate
```{r}
fit_bod_ts <- HoltWinters(bod_ts, gamma=FALSE)
fit_bod_ts
pred_bod_ts <- forecast(fit_bod_ts,3)


flag <- FALSE
for ( i in pred_bod_ts$mean)  if ( i < 0) flag <- TRUE
if (flag == TRUE)
  fit_bod_ts <- HoltWinters(bod_ts, beta = FALSE,gamma=FALSE)
  fit_bod_ts
  pred_bod_ts <- forecast(fit_bod_ts,3)
plot(pred_bod_ts)
pred_bod_ts

```

## Nitrate
N+ and N levels
```{r}
fit_nitrate_ts <- HoltWinters(nitrate_ts, gamma=FALSE)
fit_nitrate_ts
pred_nitrate_ts <- forecast(fit_nitrate_ts,3)


flag <- FALSE
for ( i in pred_nitrate_ts$mean)  if ( i < 0) flag <- TRUE
if (flag == TRUE)
  fit_nitrate_ts <- HoltWinters(nitrate_ts,beta = FALSE, gamma=FALSE)
  fit_nitrate_ts
  pred_nitrate_ts <- forecast(fit_nitrate_ts,3)
````

````
plot(pred_nitrate_ts)
pred_nitrate_ts

```


## Fecal
Level of fecal contamination
```{r}
fit_fecal_ts <- HoltWinters(fecal_ts, gamma=FALSE)
fit_fecal_ts
pred_fecal_ts <- forecast(fit_fecal_ts,3)


flag <- FALSE
for ( i in pred_fecal_ts$mean)  if ( i < 0) flag <- TRUE
if (flag == TRUE)
  fit_fecal_ts <- HoltWinters(fecal_ts, beta = FALSE, gamma=FALSE)
  fit_fecal_ts
  pred_fecal_ts <- forecast(fit_fecal_ts,3)

plot(pred_fecal_ts)
pred_fecal_ts

```
````
````