Outlier Correction

On this page

Use

Prerequisite

Settings

The outlier correction algorithm detects and corrects outliers in time series data before forecast calculations are performed.

Use

12.8

An outlier is a value in the historical data that lies outside the accepted range of values, which is also called the tolerance lane. There can be a variety of reasons for this deviation; for example, a data entry error or a one-time event that affected the sales results.

forecast algorithms are run.

You can add the algorithm to your forecast model on the **Preprocessing Steps** tab. When you do so, you can specify the detection and correction methods to be applied. You can also set your own outlier markers to override the results of automatic outlier detection.

Outliers that are detected or marked are **not** corrected in the following cases:

- When the correction method is set to **No Correction**
- When a correction method is selected but the **Consider Time Series Properties** setting is enabled, seasonality or trend is found in the data, and the correction method for trend and seasonality is set to **No Correction**

When an outlier is not corrected for the above reasons, it is still saved and displayed in the application log. It is also considered by the **gradient boosting of decision trees** algorithm if that it is set to use the **Periods with Outliers** system-generated feature.

If you want to use the same outlier detection or correction method for various purposes, you can add more than one input or output key figure to the same outlier correction algorithm. For each input, you can select a key figure that should store your outlier markers.

In addition to setting multiple key figures for the outlier correction algorithm, you can also add the algorithm multiple times to a forecast model. This is useful if you want to try different detection or correction modes on different input key figures. For example, you may want to run an interquartile range test on the values of the sales history key figure and a variance test on the delivered quantity key figure.

On this page

Use

Prerequisite

Settings

Prerequisite

Missing Values	
Missing (null) values are always replaced before forecasting activity in the planning model.	Select the value 1 for the `FORECAST_ESCAPENULL` global configuration parameter in the **Global Configuration** app.
Missing (null) values are replaced only when you select the current forecast model for a forecasting run.	Add the **substitute missing values** algorithm to your forecast model on the **Preprocessing Steps** tab.

On this page

Use

Prerequisite

Settings

Settings

You can specify the following settings for the outlier correction algorithm:

- **Outlier Detection Methods**
  - **Interquartile Range Test**

    The system checks whether the time series values are within the interquartile range, which is the difference between the third quartile and the first quartile of the data. The values that are not within this range are identified as outliers.

    This is the default outlier detection method. We recommend that you choose it if you don't want the outliers in your time series data to influence the detection results.

    In statistics, the quartiles are the three values that divide the data into four equal groups, each group comprising of 25% of observations from the data. The quartiles are determined as follows:

    - The first quartile (Q1) splits the lowest 25% of data from the highest 75% (therefore, it’s also

- The third quartile (Q3) splits the highest 25% of data from the lowest 75% (therefore, it's also called the 75th percentile).
- The interquartile range (IQR) is the difference between the third and first quartiles. (IQR = Q3 - Q1)

To detect the outliers using the interquartile range method, the system calculates a lower and an upper bound using the first (Q1) and the third (Q3) quartile:

- Lower bound = Q1 – Multiplier × IQR
- Upper bound = Q3 + Multiplier × IQR

The values that fall outside of this tolerance lane are considered as outliers. The multiplier influences the sensitivity of the outlier detection; using lower values for the multiplier will probably lead to the detection of more outliers in the sales history.

- **Variance Test**

  The system checks whether the historical values deviate from the mean by more than the standard deviation multiplied by a constant. The values showing larger deviations are identified as outliers.

  To detect the outliers using the variance test method, the system calculates a lower and an upper bound using the mean and the standard deviation (SD) of the historical data:

  - Lower bound = Mean – Multiplier × SD
  - Upper bound = Mean + Multiplier × SD

  The values that fall outside of this tolerance lane are considered as outliers. The multiplier influences the sensitivity of the outlier detection; using lower values for the multiplier will probably lead to the detection of more outliers in the sales history.

On this page

Use

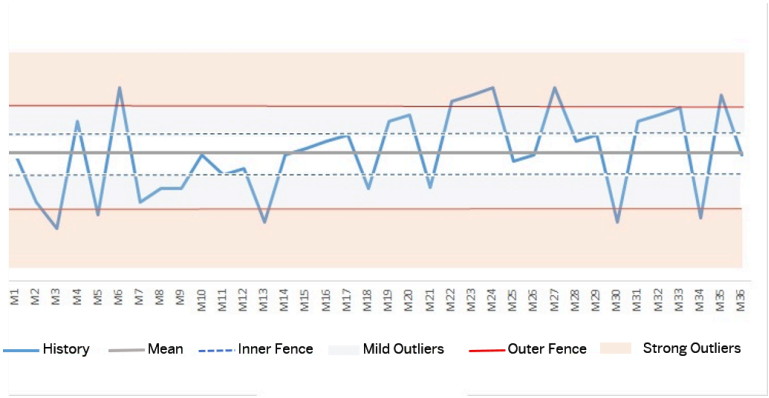
Prerequisite

Settings

additional values in the range or excluding a set of values from it. The most commonly used multipliers are 1.5 and 3.

If you choose 1.5 as the multiplier, the lower and upper bound determine a so called inner fence. The values outside the inner fence are identified as outliers but this identification sometimes brings false results. To exclude false outliers, you need to review the detection results manually, which requires a forecast model that only contains preprocessing steps.

Therefore, we recommend that you choose 3 as the multiplier. In this case, the lower and upper bound determine an outer fence. The values outside the outer fence are considered strong outliers. Strong outliers are always identified correctly so you don't need to review them manually.



- **Outlier Correction Methods**

  The following methods are available:

On this page

Use

Prerequisite

Settings

	the outliers so that they are at the limits of the new tolerance range. This is the default outlier correction method.
No Correction	Outliers are not corrected; outlier detection is used for information purposes only.
Correction with Mean	The system replaces outliers with the average of all key figure values calculated for the historical periods.
Correction with Median	The system replaces outliers with the median of all key figure values calculated for the historical periods. The outliers are excluded from the calculation.
Smoothing	The system smoothes the outliers into the mean of the time series across the period with the outlier and a given number of periods on both sides of it. The number of periods considered on each side depends on the periodicity selected for the forecast model: **Number of Periods in Smoothing Windows** **Periodicity** / **Periods in Smoothing Windows** Day / 7 Week / 3

On this page

Use

Prerequisite

Settings

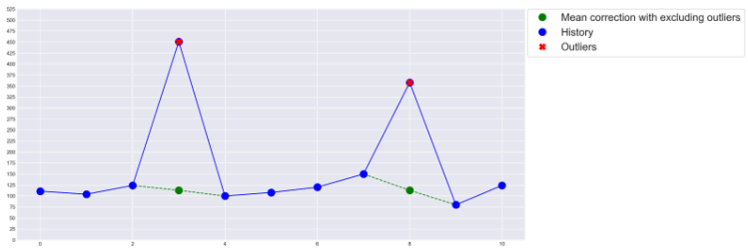
	Windows
Month	2
Quarter	2
Year	1

The way correction is performed also depends on whether the **Detect Level Shifts** option is enabled for the algorithm. For more information, see *Detect Level Shifts* below.

- **Include Outliers**

  You can use this setting to specify if the outliers should be included or excluded from the time series when the outlier correction step is performed by the selected method.

  For example, you can set the algorithm to use the **correction with mean** method and exclude the outliers when calculating the mean for the correction. The result of this setup is shown in the following chart:



  The chart below shows the corrections of the same time series when outliers are included in the calculation of the mean:

On this page

Use

Prerequisite

Settings

Was this page helpful?

> **Note**
>
> The **Include Outliers** setting is enabled by default for all available correction methods and can't be disabled if the **smoothing** method is selected for outlier correction. However, outliers that are larger than the average of their smoothing windows multiplied by 4 are automatically excluded from the calculations. This is to make sure that factors causing such significant differences are considered in forecasting.

- **Detect Level Shifts**

  If you select this option and the algorithm detects level shifts in the time series, outliers are corrected for each level separately.

  > **Note**
  >
  > Level shift detection is skipped if the **Consider Time Series Properties** option is also enabled for the forecast model.

  If the option is left disabled, the corrected values are calculated for the entire historical horizon or for each subset of smoothing periods, regardless of any level shifts that may have been detected in the time series.

  For example, if for some periods the sales values are close to 1000 and then for some periods they are around 2000, there is a level shift in the time series. When the **Smoothing** method is selected and level shift detection is enabled, outliers are smoothed into a mean of 1000 and 2000, respectively. When level shift detection is disabled, they are smoothed into a total mean value of 1500.

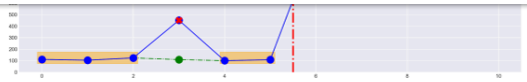
  The following chart shows forecast results calculated by a model in which level shift detection is enabled:

On this page

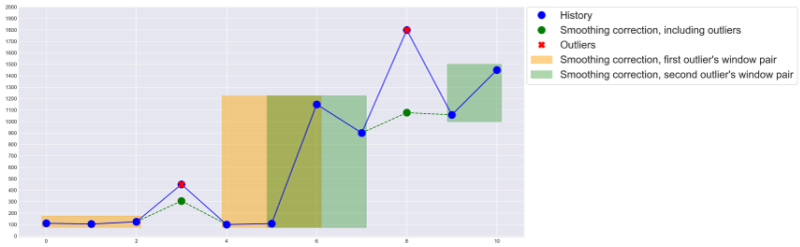
Use

Prerequisite

Settings



The chart below shows forecasts for the same time series when level shift detection is disabled:



- **Consider Time Series Properties**

  If you select this option, the algorithm will consider the results of the most recently executed time series analysis jobs. This means that it will detect and correct outliers that significantly differ from the moving average calculated for the values of any identified trend, seasonality, or both.

  > **Note**
  >
  > Time series properties can only be considered if a **Forecast Automation** application job was previously executed on the same calculation level as the one defined for the forecast model.

  Considering time series properties can help the algorithm find outliers that don't vary significantly from the mean or median but do vary from the seasonality or trend pattern in the data. It can also help prevent the algorithm from identifying values as outliers just because they are parts of a pattern. For example, the highest value in a trend may be considered an outlier and wrongly corrected with the selected method. This can be avoided if the trend property is considered by the algorithm.

- **Correction Method for Seasonality and Trend**

On this page

Use

Prerequisite

Settings

Method	Description
No Correction	Outliers are not corrected even if a general correction method is set for the algorithm. Outlier detection is used for information purposes only.
Adjustment to Seasonality and Trend	Outliers are changed to fit the moving average calculated for the values in seasonality and trend patterns.
Adjustment to Seasonality and Trend with Tolerance	The system recalculates the tolerance range to follow the values of the moving average calculated for seasonality and trend patterns, and it changes the outliers so that they are within the new tolerance range.

For more information, see Forecast Automation.

- **Smoothing Window**

  When the **Adjustment to Seasonality and Trend** or **Adjustment to Seasonality and Trend with Tolerance** method is selected and the time series shows seasonality or trend patterns, the algorithm calculates the moving average of the values that are forming these patterns. For this calculation, it needs information on the subset size, which is also known as the smoothing window.

  If trend and seasonality are both found in the time series, the algorithm uses the length of the seasonal cycle as the smoothing window. This length can be calculated by the system, or you can specify it manually in the **Manage Forecast Automation Profiles** app.

On this page

Use

Prerequisite

Settings

the periodicity chosen for your forecast model. We recommend that you define 6-12 months, 4-12 weeks, or 7-30 days for this setting.

> **Note**
>
> If the smoothing window is too short, the system may calculate too volatile trend patterns for your data. For example, if you work with daily periodicity and you have higher sales values on the weekends, these higher values may distort the moving average of the overall trend pattern. To avoid this, you should choose a smoothing window of at least 7 days.

- **Input for Algorithm**

  The key figure that the system is to examine with regards to outliers.

- **Outlier Markers**

  The key figure in which the system is to save manually defined outlier markings for any period in the historical horizon. It can store three types of markings:

  - 1, for periods in which the values that should be considered outliers and should be corrected by the selected correction method regardless of whether they were identified by the algorithm as outliers
  - 2, for periods in which the values should not be considered outliers even if they were identified by the algorithm as outlier.
  - Null or any value other than 1 or 2, for periods in which the results of automatic outlier detection should be considered

  > **Note**
  >
  > The key figure should only store integers.

- **Save Result In**

On this page

Use

Prerequisite

Settings

SAP IBP for demand doesn't support the use of external key figures or key figures with external key figures in their calculations.

On this page

Use

Prerequisite

Settings