

Forecast KPIs: RMSE, MAE, MAPE & Bias

Choosing the right forecasting metric is not straightforward. Let's review the pro and con of RMSE, MAE, MAPE, and Bias. Spoiler: MAPE is the worst. Don't use it.

This article is an extract from my book [Data Science for Supply Chain Forecasting](#). You can read my other articles [here](#). I am also active on [LinkedIn](#).



Credit

Measuring forecast accuracy (or error) is not an easy task as **there is no one-size-fits-all indicator**. Only experimentation will show you what Key Performance Indicator (KPI) is best for you. As you will see, each indicator will avoid some pitfalls but will be prone to others.

. . .

The first distinction we have to make is the difference between the precision of a forecast and its bias:

- **Bias** represents the historical average error. Basically, will your forecasts be, on average, too high (i.e., you *overshot* the demand)

or too low (i.e., you *undershot* the demand)? This will give you the overall direction of the error.

- **Precision** measures how much spread you will have between the forecast and the actual value. The precision of a forecast gives an idea of the magnitude of the errors but not their overall direction.

Of course, as you can see in the figure below, what we want to have is a forecast that is both precise and unbiased.

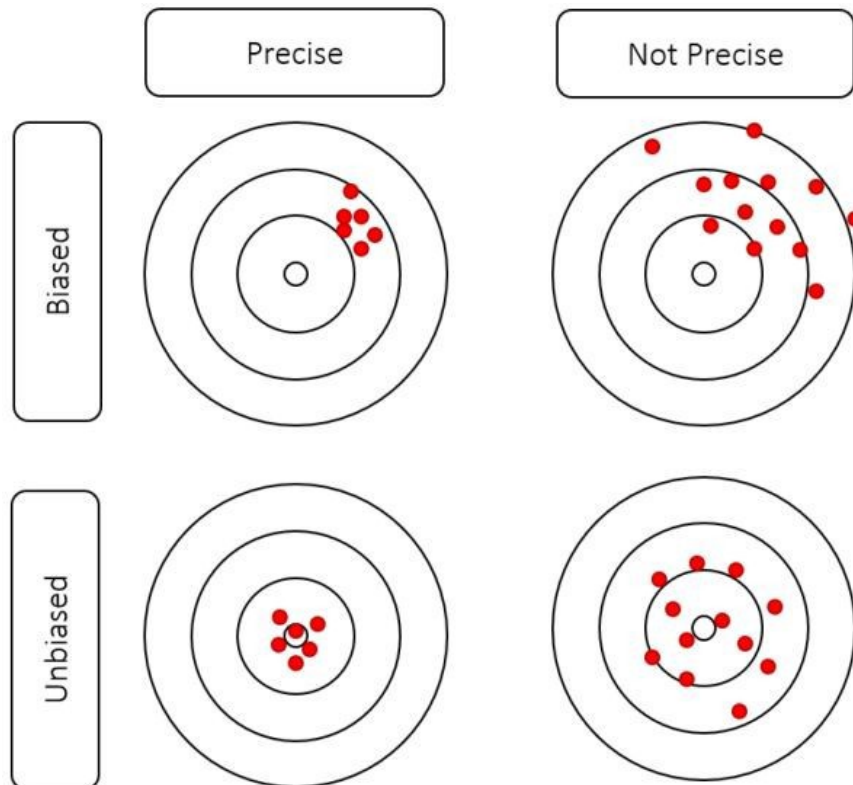


Figure 2.1: Precision & Bias

Forecast KPI

Error

Let's start by defining the error as the forecast minus the demand.

Note that if the forecast overshoots the demand with this definition, the error will be positive. If the forecast undershoots the demand, then the error will be negative.

$$e_t = f_t - d_t$$

Bias

The bias is defined as the average error:

$$bias = \frac{1}{n} \sum_n e_t$$

where n is the number of historical periods where you have both a forecast and a demand.

As a positive error on one item can offset a negative error on another item, a forecast model can achieve very low bias and not be precise at the same time. Obviously, the bias alone won't be enough to evaluate your forecast precision. But a highly biased forecast is already an indication that something is wrong in the model.

MAPE

The **Mean Absolute Percentage Error (MAPE)** is one of the most commonly used KPIs to measure forecast accuracy.

MAPE is the sum of the individual absolute errors divided by the demand (each period separately). It is the average of the percentage errors.

$$MAPE = \frac{1}{n} \sum \frac{|e_t|}{d_t}$$

MAPE is a really strange forecast KPI.

It is quite well-known among business managers, despite being a poor-accuracy indicator. As you can see in the formula, MAPE divides each error individually by the demand, so it is skewed: high errors during low-demand periods will significantly impact MAPE. Due to this,

optimizing MAPE will result in a strange forecast that will most likely undershoot the demand. Just avoid it.

MAE

The **Mean Absolute Error (MAE)** is a very good KPI to measure forecast accuracy. As the name implies, it is the mean of the absolute error.

$$MAE = \frac{1}{n} \sum |e_t|$$

One of the first issues of this KPI is that it is not scaled to the average demand. If one tells you that MAE is 10 for a particular item, you cannot know if this is good or bad. If your average demand is 1000, it is, of course, astonishing. Still, if the average demand is 1, this is a very poor accuracy. To solve this, it is common to divide MAE by the average demand to get a %:

$$MAE\% = \frac{\frac{1}{n} \sum |e_t|}{\frac{1}{n} \sum d_t} = \frac{\sum |e_t|}{\sum d_t}$$

MAPE/MAE Confusion—It seems that many practitioners use the MAE formula and call it MAPE. This can cause a lot of confusion. When discussing forecast error with someone, I would always advise you to explicitly show how you compute the forecast error to be sure to compare apples and apples.

RMSE

The **Root Mean Squared Error (RMSE)** is a strange KPI but a very helpful one, as we will discuss later. It is defined as the square root of the average squared error.

$$RMSE = \sqrt{\frac{1}{n} \sum e_t^2}$$

Just as for MAE, RMSE is not scaled to the demand. We can then define RMSE% as such,

$$RMSE\% = \frac{\sqrt{\frac{1}{n} \sum e_t^2}}{\frac{\sum d}{n}}$$

Actually, many algorithms (especially for machine learning) are based on the **Mean Squared Error (MSE)**, which is directly related to RMSE.

$$MSE = \frac{1}{n} \sum e_t^2$$

Many algorithms use MSE as it is faster to compute and easier to manipulate than RMSE. But it is not scaled to the original error (as the error is squared), resulting in a KPI that we cannot relate to the original demand scale. Therefore, we won't use it to evaluate our statistical forecast models.

A question of error weighting

Compared to MAE, RMSE does not treat each error the same. It gives more importance to the most significant errors. That means that one big error is enough to get a very bad RMSE.

Let's do an example with a dummy demand time series.

Period	1	2	3	4	5	6	7	8	9	10	11	12
Demand	10	12	14	8	9	5	8	10	12	11	10	15

Let's imagine we want to compare two slightly different forecasts. The only difference is the forecast on the latest demand observation: forecast #1 undershot it by 7 units and forecast #2 by *only* 6 units.

Period	1	2	3	4	5	6	7	8	9	10	11	12
Demand	10	12	14	8	9	5	8	10	12	11	10	15
Forecast #1	12	14	15	10	7	4	5	8	12	14	13	8
Error #1	2	2	1	2	-2	-1	-3	-2	0	3	3	-7
Forecast #2	12	14	15	10	7	4	5	8	12	14	13	9
Error #2	2	2	1	2	-2	-1	-3	-2	0	3	3	-6

If we look at the KPI of these two forecasts, this is what we obtain:

KPI	MAE	RMSE
Forecast #1	2.33	2.86
Forecast #2	2.25	2.66

Interestingly, by just changing the error of this last period by a single unit, we decrease the total RMSE by 6.9% (2.86 to 2.66). Still, MAE is only reduced by 3.6% (2.33 to 2.25), so the impact on MAE is nearly twice as low. Clearly, RMSE emphasizes the most significant errors, whereas MAE gives the same importance to each error. You can try this for yourself and reduce the error of one of the most accurate periods to observe the impact on MAE and RMSE.

Spoiler: nearly no impact on RMSE.

As you will see later, RMSE has some other very interesting properties.

What would you like to predict?

We went through the definition of these KPIs (bias, MAPE, MAE, RMSE), but it is still unclear what difference it can make for our model to use one instead of another. One could think that using RMSE instead of MAE or MAE instead of MAPE doesn't change anything. But nothing is less true.

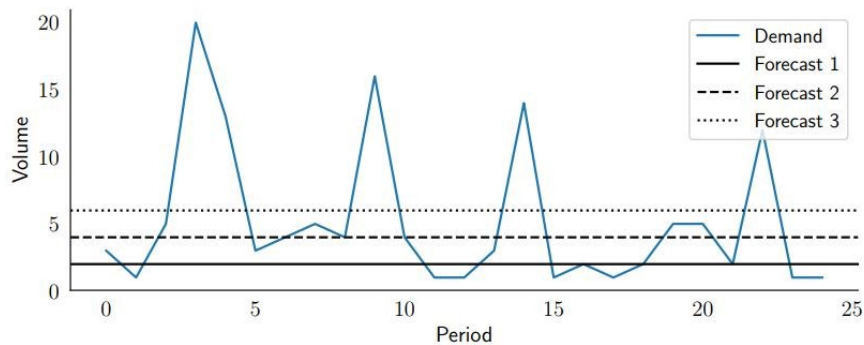
Let's do a quick example to show this.

Let's imagine a product with a low and rather flat weekly demand that has from time to time a big order (maybe due to promotions or to

clients ordering in batches). Here is the demand per week that we observed so far:

	W1	W2	W3	W4	W5
Mon	3	3	4	1	5
Tue	1	4	1	2	2
Wed	5	5	1	1	12
Thu	20	4	3	2	1
Fri	13	16	14	5	20

Now let's imagine we propose three different forecasts for this product. The first one predicts 2 pieces/day, the second one 4, and the last one 6. Let's plot the demand we observed and these forecasts.



Let's see how each of these forecasts performs in terms of bias, MAPE, MAE, and RMSE on the historical period:

	Forecast 1	Forecast 2	Forecast 3
Bias	-3.9	-1.9	0.1
MAPE	64%	109%	180%
MAE	4.4	4.1	4.8
RMSE	7.1	6.2	5.9

It means that forecast #1 was the best during the historical period in terms of MAPE, forecast #2 was the best in terms of MAE. Forecast #3 was the best in terms of RMSE and bias (but the worst on MAE and MAPE). Let's now reveal how these forecasts were made:

- **Forecast 1** is just a very low amount.
- **Forecast 2** is the demand median: 4.
- **Forecast 3** is the average demand.

Median vs. Average—mathematical optimization

Before discussing the different forecast KPIs further, let's take some time to understand why a forecast of the median will get a good **MAE** and a forecast of the mean a good **RMSE**.

There is a bit of math ahead. If these equations are unclear to you, this is not an issue—don't get discouraged. Just skip them and jump to the conclusion of the RMSE and MAE paragraphs.

RMSE

Let's start with RMSE:

$$RMSE = \sqrt{\frac{1}{n} \sum e_t^2}$$

To simplify the following algebra, let's use a simplified version: the Mean Squared Error (MSE):

$$MSE = \frac{1}{n} \sum e_t^2$$

If you set MSE as a target for your forecast model, it will minimize it. One can minimize a mathematical function by setting its derivative to zero. Let's try this.

$$\frac{\partial MSE}{\partial f} = \frac{\partial \frac{1}{n} \sum (f_t - d_t)^2}{\partial f}$$

$$\frac{2}{n} \sum (f_t - d_t) = 0$$

$$\sum f_t = \sum d_t$$

Conclusion to optimize a forecast's MSE, the model will have to aim for the total forecast to be equal to the total demand. That is to say that optimizing MSE aims to produce a prediction that is correct *on average* and, therefore, unbiased.

MAE

Now let's do the same for MAE.

$$\frac{\partial MAE}{\partial f} = \frac{\partial \frac{1}{n} \sum |f_t - d_t|}{\partial f}$$

Or,

$$|f_t - d_t| = \begin{cases} f_t - d_t & d_t < f_t \\ \text{indefinite} & d_t = f_t \\ d_t - f_t & d_t > f_t \end{cases}$$

and

$$\frac{\partial |f_t - d_t|}{\partial f} = \begin{cases} 1 & d_t < f_t \\ \text{indefinite} & d_t = f_t \\ -1 & d_t > f_t \end{cases}$$

Which means that

$$\frac{\partial MAE}{\partial f} = \frac{1}{n} \sum \begin{cases} 1 & d_t < f_t \\ -1 & d_t > f_t \end{cases}$$

Conclusion to optimize MAE (i.e., set its derivative to 0), the forecast needs to be as many times higher than the demand as it is lower than the demand. In other words, we are looking for a value that splits our dataset into two equal parts. This is the exact definition of the median.

MAPE

Unfortunately, the derivative of MAPE won't show some elegant and straightforward property. We can simply say that MAPE promotes a very low forecast as it allocates a high weight to forecast errors when the demand is low.

Conclusion

As we saw above, in any model, the optimization of RMSE will seek to be correct on average. In contrast, MAE's optimization will try to be as often overshooting the demand as undershooting the demand, which means targeting the demand median. We have to understand that a significant difference lies in the mathematical roots of MAE & RMSE.

One aims at the median, the second aims at the average.

MAE or RMSE— which one to choose?

Is it worse to aim for the median or the average of the demand? Well, the answer is not black and white. Each technique has some benefits and some risks, as we will discuss in the next pages. Only experimentation will reveal which method works best for a dataset. You can even choose to use both RMSE & MAE.

Let's take some time to discuss the impact of choosing either RMSE or MAE on bias, sensitivity to outliers, and intermittent demand.

Bias

For many products, you will observe that the median is not the same as the average demand. The demand will most likely have some peaks here and there that will result in a skewed distribution. These skewed

demand distributions are widespread in supply chains as the peaks can be due to periodic promotions or clients ordering in bulk. This will cause the demand median to be below the average demand, as shown below.

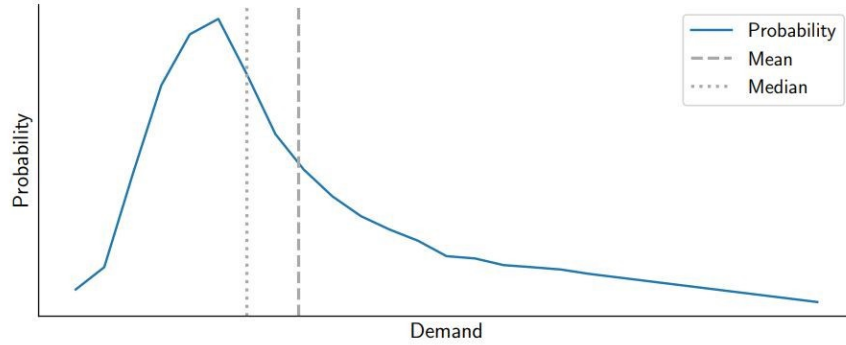


Figure 2.2: Median vs Average

This means that a forecast that is minimizing MAE will result in a bias. In comparison, a forecast minimizing RMSE will not result in bias (as it aims for the average). This is definitely MAE’s main weakness.

Sensitivity to outliers

As we discussed, RMSE gives greater importance to the highest errors. This comes at a cost: a sensitivity to outliers. Let’s imagine an item with the following demand pattern.

The median is 8.5, and the average is 9.5. We already observed that if we make a forecast that minimizes MAE, we will forecast the median (8.5), and we would be, on average undershooting the demand by 1 unit (bias = -1). You might then prefer to minimize RMSE and to forecast the average (9.5) to avoid this situation.

Nevertheless, let’s now imagine that we have one new demand observation of 100.

Period	1	2	3	4	5	6	7	8	9	10
Demand	16	8	12	9	6	12	5	7	6	100

The median is still 8.5 (it hasn’t changed!), but the average is now 18.1. In this case, you might not want to forecast the average and might revert back to a forecast of the median.

Generally speaking, the median is more robust to outliers than the average. This is rather important in a supply chain environment as we can face many outliers due to encoding mistakes or demand peaks (marketing, promotions, spot deals).

Is robustness to outliers always a good thing? No.

Intermittent demand

Indeed, unfortunately, the median's robustness to outliers can result in a very annoying effect for items with intermittent demand.

Let's imagine that we sell a product to a single client. It is a highly profitable product. Unfortunately, our unique client seems to make an order one week out of three without any kind of pattern. The client always orders the product in batches of 100. We then have an average weekly demand of 33 pieces and a demand median of... 0.

We have to populate a weekly forecast for this product. Let's imagine we do a first forecast that aims for the average demand (33 pieces). Over the long-term, we will obtain a total squared error of 6,667 (RMSE of 47) and a total absolute error of 133 (MAE of 44).

Week	Demand	Forecast	Error	Error	Error ²
1	100	33	67	67	4 445
2	0	33	-33	33	1 111
3	0	33	-33	33	1 111
Total	100	100	0	133	6 667

Now, if we forecast the demand median (0), we obtain a total absolute error of 100 (MAE of 33) and a total squared error of 10.000 (RMSE of 58).

Week	Demand	Forecast	Error	Error	Error ²
1	100	0	-100	100	10 000
2	0	0	0	0	0
3	0	0	0	0	0
Total	100	0	-100	100	10 000

As we can see, MAE is a bad KPI to use for intermittent demand. As soon as you have more than half of the periods without demand, the optimal forecast is... 0!

Conclusion

MAE protects outliers, whereas RMSE assures us to get an unbiased forecast. Which indicator should you use? There is, unfortunately, no definitive answer. As a supply chain data scientist, you should experiment: if using MAE as a KPI results in a high bias, you might want to use RMSE. If the dataset contains many outliers, resulting in a skewed forecast, you might want to use MAE.

Note that you can choose to report forecast error with one or more KPIs (typically MAE & bias) and use another one (RMSE?) to optimize your models.

The last trick to use against low-demand items is to aggregate the demand to a higher time horizon. For example, if the demand is low at a weekly level, you could test a monthly forecast or even a quarterly forecast. You can always disaggregate the forecast back into the original time bucket by simply dividing it. This technique can allow you to use MAE as a KPI and smooth demand peaks simultaneously.

About the author

Follow me on LinkedIn!

Nicolas Vandepuut — Consultant, Founder —
SupChains | LinkedIn

View Nicolas Vandepuut's profile on LinkedIn, the
world's largest professional community. Nicolas ...

www.linkedin.com



Nicolas Vandepuut is a supply chain data scientist specialized in demand forecasting and inventory optimization. He founded his consultancy company SupChains in 2016 and co-founded SKU Science —a fast, simple, and affordable demand forecasting platform—in 2018. Passionate about education, Nicolas is both an avid learner and enjoys teaching at universities: he has taught forecasting and inventory optimization to master students since 2014 in Brussels, Belgium. Since 2020, he is also teaching both subjects at CentraleSupélec, Paris, France. He published *Data Science for Supply Chain Forecasting* in 2018 (2nd edition in 2021) and *Inventory Optimization: Models and Simulations* in 2020.