How to Estimate Forecasting Quality: A System-Motivated Derivation of Symmetric Mean Absolute Percentage Error (SMAPE) and Other Similar Characteristics

Vladik Kreinovich[1], Hung T. Nguyen[2,3], and Rujira Ouncharoen[4]
[1]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu
[2]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA, hunguyen@nmsu.edu
[3]Faculty of Economics
[4]Department of Mathematics
Chiang Mai University, Thailand, rujira@chiangmai.ac.th

Abstract

When comparing how well different algorithms forecast time series, researchers use an average value of the ratio $\frac{|x-y|}{(|x|+|y|)/2}$, known as the Symmetric Mean Absolute Percentage Error (SMAPE). In this paper, we provide a system-motivated explanation for this formula. We also explain how this formula explains the use of geometric mean to combine different forecasts.

1 Introduction: What Is SMAPE

Forecasting is important. One of the main problems of science is predicting (forecasting) future events: we need to predict tomorrow's weather, we need to predict future locations of celestial bodies, we would like to be able to predict earthquakes, etc.

An important particular case of this problem is forecasting time series, when we use the previous values of a physical quantity to predict its future values; see, e.g., [2, 4, 5, 6, 7, 8, 9, 10, 11].

Need to compare forecasting methods. Many researchers are working on time series forecasting. There are many software packages for forecasting, there are regular competitions whose aim is to decide which packages forecast better. How do we make this decision?

How to compare: general idea. In forecasting competitions, we deal with the sequentially measured values $x(t)$ of some physical quantity. For each time series, all the competing packages take the initial part $x(1), \ldots, x(t_0)$ of the resulting time series as an input, and produce predictions $y(t_0+1), y(t_0+2), \ldots$ for the future values $x(t_0+1), x(t_0+2), \ldots$

We then select a measure of similarity $s(x, y)$ and use the sum

$$\sum_k s(x(t_0+k), y(t_0+k))$$

to gauge the forecasting quality. The sum of the sums corresponding to different time series is then used to gauge the quality of a forecasting algorithm.

Selecting measure of similarity. Which measure of similarity $s(x, y)$ should we select?

Historically first choice: least squares. The first forecasting competitions used the expression $s(x, y) = (x-y)^2$ for which the resulting sum

$$\sum_k s(x(t_0+k), y(t_0+k)) = \sum_k (x(t_0+k) - y(t_0+k))^2$$

corresponds to the usual least-squares accuracy measure.

Limitation of least squares. The problem with this measure is that different time series have different numerical scales. As a result, the overall performance measure was almost completely determined by a few time series with the largest values of $x(t)$ – for which the differences $x - y$ are the largest; see, e.g., [10].

From absolute to relative approximation errors: a way to avoid the above limitation. The original similarity measure $s(x, y) = (x-y)^2$ corresponds to the *absolute* approximation error $|x-y|$: $s(x, y) = |x-y|^2$. To avoid this dependence on scale, researchers proposed to use scale-invariant characteristics such as the *relative* approximation error $s(x, y) = \dfrac{|x-y|}{|x|}$.

Limitation of the relative error. The problem with the relative error is that for a possible case when the actually measured value x is 0, this measure becomes infinite – and thus, the sum characterizing each forecasting method becomes infinite – which is useless for comparison; see, e.g., [10].

Enter SMAPE. To avoid the above problem, researchers have proposed to modify relative error, by replacing $|x|$ in the denominator with an arithmetic average. The resulting similarity measure

$$s(x, y) = \frac{|x-y|}{(|x|+|y|)/2} \tag{1}$$

is known as the *Symmetric Mean Absolute Performance Error* (SMAPE, for short).

This is the main similarity measure used in the time series forecasting competitions.

Natural question. A natural question is: why divide by arithmetic average? Why, e.g., not by $\max(|x|, |y|)$? Why not by some other similar function? All these functions would retain scale-invariance and avoid infinities, so why SMAPE?

What we do in this paper. In this paper, we show that for non-negative values, the idea of relative error naturally leads to SMAPE. This explains – at least for non-negative values – why SMAPE is efficiently used.

We also explain how SMAPE leads to the fact that geometric-mean combination of different forecasts sometimes results in a better forecast; see, e.g., [3].

After that, we analyze similar problems and describe corresponding similarity measures.

2 Why SMAPE: A System-Motivated Explanation

Analysis of measurement and approximation accuracy: an important part of systems approach. Our objective is to gauge approximation accuracy. In general, taking into account measurement and approximation accuracy is an important aspect of the system approach; see, e.g., [1]. It is therefore reasonable to use this approach as a motivation in our search for an explanation of why SMAPE is used.

In the traditional analysis of measurement and approximation errors, usually two types of errors are considered:

- *absolute* errors, i.e., errors whose bound δ does not depend on the actual value v of the measured quantity, and
- *relative* errors, i.e., errors whose bound growth linearly with the actual value v of the measured quantity, as $\varepsilon \cdot v$.

Let us use this distinction when analyzing our problem.

Relative error: towards a formal definition. Let us assume that the actual value of some quantity is v, and let $\varepsilon > 0$ by a real number. We say that an estimate x for v is accurate with relative accuracy ε if $v \cdot (1-\varepsilon) \leq x \leq v \cdot (1+\varepsilon)$. In precise terms, this can be described as follows.

Definition 1. *We say that a number x* approximates *a number v* with relative accuracy *$\varepsilon \geq 0$ if $v \cdot (1-\varepsilon) \leq x \leq v \cdot (1+\varepsilon)$.*

Analysis of the problem. For each moment t, we have two approximate numbers corresponding to this moment of time:

- the value $x(t)$ resulting from the measurement, and
- the value $y(t)$ resulting from the forecast.

Both values are approximations to the actual (unknown) value of the corresponding quantity at this moment of time:

- the measurement result is approximate since measurements are never absolutely accurate, and
- the forecasting result is approximate since forecasting is never absolutely accurate.

It is therefore reasonable to gauge the closeness of two estimates x and y as the smallest $\varepsilon \ge 0$ for which both estimates approximate some value v with relative accuracy ε.

Definition 2. *By the* relative measure of similarity $R(x, y)$ *between two non-negative real numbers x and y, we mean the smallest possible $\varepsilon \ge 0$ for which both x and y approximate some real number v with relative accuracy ε.*

Proposition 1. $R(x, y) = \dfrac{|x - y|}{x + y}$.

Comment. This result explains the SMAPE formula.

Proof. Without losing generality, let us assume that $x \le y$. We want to find the smallest $\varepsilon \ge 0$ for which there exists a v whom both x and y approximate with this relative accuracy, i.e., for which the following two double inequalities are consistent:

$$v \cdot (1 - \varepsilon) \le x \le v \cdot (1 + \varepsilon); \tag{2a}$$

$$v \cdot (1 - \varepsilon) \le y \le v \cdot (1 + \varepsilon). \tag{2b}$$

The desired smallest value ε is always smaller than or equal to 1, since for $\varepsilon = 1$, we can take $v = y$. Thus, $\varepsilon \le 1$ and $1 - \varepsilon \ge 0$.

Dividing both sides of the inequality $v \cdot (1 - \varepsilon) \le x$ by $1 - \varepsilon$, we transform it into an equivalent form $v \le \dfrac{x}{1 - \varepsilon}$. Similarly, all inequalities (2) can be equivalently transformed into inequalities directly binding v:

$$v \le \frac{x}{1 - \varepsilon}; \tag{3aa}$$

$$\frac{x}{1 + \varepsilon} \le v; \tag{3ab}$$

$$v \le \frac{y}{1 - \varepsilon}; \tag{3ba}$$

$$\frac{y}{1 + \varepsilon} \le v. \tag{3bb}$$

Since $x \le y$, the inequality (3ba) follows from (3aa), and the inequality (3ab) follows from (3bb). Thus, for all four inequalities to be satisfied, it is sufficient that the inequalities (3aa) and (3bb) be satisfied, i.e., that

$$\frac{y}{1+\varepsilon} \le v \le \frac{x}{1-\varepsilon}. \tag{4}$$

In other words, we have a lower bound and an upper bound for the desired value v. The existence of such v is equivalent to the fact that the lower bound is smaller than or equal to the upper bound, i.e., that

$$\frac{y}{1+\varepsilon} \le \frac{x}{1-\varepsilon}. \tag{5}$$

Multiplying both sides of this inequality by $1+\varepsilon$ and $1-\varepsilon > 0$, we get an equivalent inequality

$$y \cdot (1-\varepsilon) \le x \cdot (1+\varepsilon), \tag{6}$$

i.e.,

$$y - \varepsilon \cdot y \le x + \varepsilon \cdot x. \tag{7}$$

By moving all the terms containing ε to the right-hand side and all other terms to the left-hand side, we get

$$y - x \le (x+y) \cdot \varepsilon. \tag{8}$$

Dividing both sides of the inequality (8) by $y+x$, we get an equivalent inequality

$$\frac{y-x}{x+y} \le \varepsilon. \tag{9}$$

The smallest value ε that satisfies this inequality is $\varepsilon = \dfrac{y-x}{x+y}$.

The proposition is proven.

3 Why Geometric Mean? A SMAPE-Based Explanation

Empirical fact. If we have several different forecasts, then combining them can lead to a better forecast. Often, the most accurate combination is a geometric mean.

In this section, we use SMAPE to explain this empirical fact.

Formulation of the problem. Suppose that we have two values x and y, we need to replace both with a single estimate z. According to the the SMAPE approach, we must select z for which the sum

$$s(x,z) + s(y,z) = \frac{|x-z|}{x+z} + \frac{|y-z|}{y+z} \tag{10}$$

is the smallest possible.

Proposition 2. *For every x and y, the smallest possible value of the sum* (10) *is attained when $z = \sqrt{x \cdot y}$.*

Comment. This result explains the use of geometric mean.

Proof. One can check that the minimum is attained when z is between x and y: otherwise we can move z closer to the interval and decrease the sum. Without losing generality, we can assume that $x \le y$. Under this assumption, the expression (10) takes the form

$$\frac{z-x}{x+z} + \frac{y-z}{y+z}. \tag{11}$$

Differentiating this expression with respect to z and equating this derivative to 0, we get

$$\frac{1 \cdot (z+x) - 1 \cdot (z-x)}{(x+z)^2} + \frac{(-1) \cdot (y+z) - 1 \cdot (y-z)}{(y+z)^2} = 0, \tag{12}$$

i.e.,

$$\frac{2x}{(x+z)^2} = \frac{2y}{(y+z)^2}. \tag{13}$$

Multiplying both sides by both denominators and dividing both sides by 2, we conclude that

$$x \cdot (y+z)^2 = y \cdot (x+z)^2, \tag{14}$$

i.e., opening the parentheses,

$$x \cdot y^2 + 2 \cdot x \cdot y \cdot z + x \cdot z^2 = y \cdot x^2 + 2 \cdot y \cdot x \cdot z + y \cdot z^2. \tag{15}$$

If we cancel equal terms in both sides, move terms proportional to z^2 to the right-hand side and all other terms to the left-hand side, we get

$$x \cdot y^2 - y \cdot x^2 = z^2 \cdot (y-x). \tag{16}$$

Here, $x \cdot y^2 - y \cdot x^2 = x \cdot y \cdot (y-x)$. Dividing both sides of the equality (16) by $y-x$, we get $x \cdot y = z^2$, i.e., $z = \sqrt{x \cdot y}$, which is exactly the geometric average.

The proposition is proven.

4 Beyond SMAPE: What Happens in Similar Situations

4.1 What if we use absolute errors instead of relative errors?

Natural question. The first natural question is: what happens if we use absolute approximation errors instead of relative ones? Let us formulate this problem in precise terms.

Definition 3. *We say that a number x* approximates *a number v* with absolute accuracy $\delta \geq 0$ *if* $v - \varepsilon \leq x \leq v + \delta$.

Definition 4. *By the* absolute measure of similarity $A(x, y)$ *between two real numbers x and y, we mean the smallest possible $\delta \geq 0$ for which both x and y approximate some real number v with absolute accuracy δ.*

Proposition 3. $A(x, y) = \dfrac{|x - y|}{2}$.

Proof. Without losing generality, let us assume that $x \leq y$. We want to find the smallest $\delta \geq 0$ for which there exists a v whom both x and y approximate with this absolute accuracy, i.e., for which the following two double inequalities are consistent:

$$v - \delta \leq x \leq v + \delta; \tag{17a}$$

$$v - \delta \leq y \leq v + \delta. \tag{17b}$$

Moving δ to the other side of the inequality $v - \delta \leq x$, we transform it into an equivalent form $v \leq x + \delta$. Similarly, all inequalities (10) can be equivalently transformed into inequalities directly binding v:

$$v \leq x + \delta; \tag{18aa}$$

$$x - \delta \leq v; \tag{18ab}$$

$$v \leq y + \delta; \tag{18ba}$$

$$y - \delta \leq v. \tag{18bb}$$

Since $x \leq y$, the inequality (18ba) follows from (18aa), and the inequality (18ab) follows from (18bb). Thus, for all four inequalities to be satisfied, it is sufficient that the inequalities (18aa) and (18bb) be satisfied, i.e., that

$$y + \delta \leq v \leq x + \delta. \tag{19}$$

Here, we have a lower bound and an upper bound for the desired value v. The existence of such v is equivalent to the fact that the lower bound is smaller than or equal to the upper bound, i.e., that

$$y - \delta \leq x + \delta. \tag{20}$$

By moving all the terms containing δ to the right-hand side and all other terms to the left-hand side, we get

$$y - x \leq 2\delta. \tag{21}$$

Dividing both sides of the inequality (21) by 2, we get an equivalent inequality

$$\frac{y - x}{2} \leq \delta. \tag{22}$$

The smallest value δ that satisfies this inequality is $\delta = \dfrac{y - x}{2}$.

The proposition is proven.

4.2 What if we have both relative and absolute errors

Question. What if we have both relative and absolute errors, with the absolute errors bounded by some constant $\delta > 0$? In this case, we have the following result.

Definition 5. *We say that a number x* approximates *a number v* with relative accuracy $\varepsilon \geq 0$ *and absolute accuracy $\delta > 0$ if*

$$v \cdot (1 - \varepsilon) - \delta \leq x \leq v \cdot (1 + \varepsilon) + \delta. \tag{23}$$

Analysis of the problem. According to Proposition 3, when $|x - y| \leq 2\delta$, there is no need to invoke relative error: the difference between x and y can be explained by using only the absolute error of given size. When $|x - y| > 2\delta$, this explanation is no longer possible, so we need to invoke relative error $\varepsilon > 0$ as well. It is reasonable to select the representation with the smallest possible relative error $\varepsilon > 0$.

Definition 6. *For each $\delta > 0$, by the δ-*relative measure of similarity $R_\delta(x, y)$ *between two non-negative real numbers x and y, we mean the smallest possible $\varepsilon \geq 0$ for which both x and y approximate some real number v with relative accuracy ε and absolute accuracy $\delta > 0$.*

Proposition 4. $R_\delta(x, y) = \dfrac{\max(|x - y| - 2\delta, 0)}{x + y}$.

Proof. Without losing generality, let us assume that $x < y$, so $y - x > 2\delta$. We want to find the smallest $\varepsilon \geq 0$ for which there exists a v whom both x and y approximate with this relative accuracy and absolute accuracy δ, i.e., for which the following two double inequalities are consistent:

$$v \cdot (1 - \varepsilon) - \delta \leq x \leq v \cdot (1 + \varepsilon) + \delta; \tag{24a}$$

$$v \cdot (1 - \varepsilon) - \delta \leq y \leq v \cdot (1 + \varepsilon) + \delta. \tag{24b}$$

Moving δ to the right-hand side of the inequality $v \cdot (1 - \varepsilon) - \delta \leq x$ and dividing both sides of the resulting inequality $v \cdot (1 - \varepsilon) \leq x + \delta$ by $1 - \varepsilon$, we transform it into an equivalent form $v \leq \dfrac{x + \delta}{1 - \varepsilon}$. Similarly, all inequalities (2) can be equivalently transformed into inequalities directly binding v:

$$v \leq \frac{x + \delta}{1 - \varepsilon}; \tag{25aa}$$

$$\frac{x - \delta}{1 + \varepsilon} \leq v; \tag{25ab}$$

$$v \leq \frac{y + \delta}{1 - \varepsilon}; \tag{25ba}$$

$$\frac{y-\delta}{1+\varepsilon} \le v. \tag{25bb}$$

Since $x \le y$, the inequality (25ba) follows from (25aa), and the inequality (25ab) follows from (25bb). Thus, for all four inequalities to be satisfied, it is sufficient that the inequalities (25aa) and (25bb) be satisfied, i.e., that

$$\frac{y-\delta}{1+\varepsilon} \le v \le \frac{x+\delta}{1-\varepsilon}. \tag{26}$$

We have a lower bound and an upper bound for the desired value v. The existence of such v is equivalent to the fact that the lower bound is smaller than or equal to the upper bound, i.e., that

$$\frac{y-\delta}{1+\varepsilon} \le \frac{x+\delta}{1-\varepsilon}. \tag{27}$$

The desired smallest value ε is always smaller than or equal to 1, since for $\varepsilon = 1$, we can take $v = y$. Thus, $\varepsilon \le 1$ and $1-\varepsilon \ge 0$.

Multiplying both sides of this inequality by $1+\varepsilon$ and $1-\varepsilon$, we get an equivalent inequality

$$(y-\delta)\cdot(1-\varepsilon) \le (x+\delta)\cdot(1+\varepsilon), \tag{28}$$

i.e.,

$$(y-\delta) - \varepsilon\cdot(y-\delta) \le (x+\delta) + \varepsilon\cdot(x+\delta). \tag{29}$$

By moving all the terms containing ε to the right-hand side and all other terms to the left-hand side, we get

$$y - x - 2\delta \le (x+y)\cdot\varepsilon. \tag{30}$$

Dividing both sides of the inequality (30) by $y+x$, we get an equivalent inequality

$$\frac{y-x-2\delta}{x+y} \le \varepsilon. \tag{31}$$

The smallest value ε that satisfies this inequality is $\varepsilon = \dfrac{y-x-2\delta}{x+y}$.

The proposition is proven.

4.3 What if relative error is fixed and we minimize absolute errors

Question. What if we fix the relative error, at some $\varepsilon > 0$, and look for the smallest possible absolute error?

Definition 7. *For each $\varepsilon > 0$, by the* ε-absolute measure of similarity $A_\varepsilon(x, y)$ *between two non-negative real numbers x and y, we mean the smallest possible $\delta \ge 0$ for which both x and y approximate some real number v with relative accuracy ε and absolute accuracy δ.*

Proposition 5. $A_\varepsilon(x,y) = \dfrac{1}{2} \cdot \max(|x-y| - \varepsilon \cdot (x+y), 0)$.

Proof. Without losing generality, let us assume that $x \le y$, and $R(x,y) > \varepsilon$, i.e., the difference between x and y cannot be explained by just a relative error ε. We want to find the smallest $\delta \ge 0$ for which there exists a v whom both x and y approximate with given relative accuracy ε and absolute accuracy δ, i.e., for which the double inequalities (24) are consistent.

We have already shown, in the proof of Proposition 4, that the existence of such v is equivalent to the inequality

$$(y-\delta) - \varepsilon \cdot (y-\delta) \le (x+\delta) + \varepsilon \cdot (x+\delta). \tag{29}$$

By moving all the terms containing δ to the right-hand side and all other terms to the left-hand side, we get

$$y - x - (x+y) \cdot \varepsilon \le 2\delta. \tag{33}$$

Dividing both sides of the inequality (33) by 2, we get an equivalent inequality

$$\frac{1}{2} \cdot (y - x - (x+y) \cdot \varepsilon) \le \delta. \tag{34}$$

The smallest value δ that satisfies this inequality is $\dfrac{1}{2} \cdot \max(|x-y| - \varepsilon \cdot (x+y), 0)$.

The proposition is proven.

Acknowledgments. This work was supported in part by a grant from Chhiang Mai University, and by the US National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

The authors are thankful to Martin Stepnicka for valuable discussions.

References

[1] A. C. Antoulas, *Approximations of Large-Scale Dynamical Systems*, SIAM, Philadelphia, Pennsylvania, 2009.

[2] J. S. Armstrong (ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers, Norwell, Massachusetts, 2001.

[3] J. S. Armstrong, "Combining forecasts", in [2].

[4] J. S. Armstrong and F. Collopy, "Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons", *International Journal of Forecasting*, 1992, Vol. 8, pp. 69–80.

[5] G. E. P. Box, G. M. Jenkins, and G. D. Reinsel, *Time Series Analysis, Forecasting, and Control*, Prentice Hall, Englewood Cliffs, New Jersey, 2008.

[6] J. C. Brocklebank and D. A. Dickey, *SAS for Forecasting Time Series*, SAS Institute, 2003.

[7] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Foirecasting*, Springer Verlag, New York, 2010.

[8] C. Chatfield, *Time-Series Foreasting*, Chapman and Hall, 2000.

[9] B. G. Flores, "A pragmatic view of accuracy measurement in forecasting", *Omega (Oxford)*, 1986, Vol. 19, No. 2, pp. 93–98.

[10] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy", *International Journal of Forecasting*, 2006, Vol. 22, No. 4, pp. 679–688.

[11] G. Shmueli, *Practical Time-Series Forecasting: A Hands-On Guide*, Statistics.com, 2014.