University of Texas at El Paso

ScholarWorks@UTEP

Departmental Technical Reports (CS)   Computer Science

9-1-2021

Why Normalized Difference Vegetation Index (NDVI)?

Francisco Zapata
The University of Texas at El Paso, fcozpt@outlook.com

Eric Smith
The University of Texas at El Paso, esmith2@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Nguyen Hoang Phuong
Thang Long University, nhphuong2008@gmail.com

Follow this and additional works at: https://scholarworks.utep.edu/cs_techrep

Part of the Computer Sciences Commons, and the Mathematics Commons
Comments:

Technical Report: UTEP-CS-21-78

Recommended Citation

Why Normalized Difference Vegetation Index (NDVI)?

Francisco Zapata, Eric Smith, Vladik Kreinovich, and Nguyen Hoang Phuong

Abstract Plants play a very important role in ecological systems – they transform CO_2 into oxygen. It is therefore very important to be able to estimate the overall amount of live green vegetation in a given area. The most efficient way to provide such a global analysis is to use remote sensing, i.e., multi-spectral photos taken from satellites, drones, planes, etc. At present, one of the most efficient ways to detect, based on remote sensing data, how much live green vegetation an area contains is to compute the value of the normalized difference vegetation index (NDVI). In this paper, we provide a theoretical explanation of why this particular index is efficient.

1 Formulation of the Problem

Empirical fact. Plants play a very important role in ecological systems – they transform CO_2 into oxygen. It is therefore very important to be able to estimate the overall amount of live green vegetation in a given area. The most efficient way to provide such a global analysis is to use remote sensing, i.e., multi-spectral photos taken from satellites, drones, planes, etc.

At present, one of the most efficient ways to detect, based on remote sensing data, how much live green vegetation an area contains is to compute the value of the normalized difference vegetation index (NDVI):

$$\text{NDVI} \overset{\text{def}}{=} \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}},$$

Francisco Zapata, Eric Smith, and Vladik Kreinovich
University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA
e-mail: fcozpt@outlook.com, esmith2@utep.edu, vladik@utep.edu

Nguyen Hoang Phuong
Division Informatics, Math-Informatics Faculty, Thang Long University, Nghiem Xuan Yem Road
Hoang Mai District, Hanoi, Vietnam, e-mail: nhphuong2008@gmail.com

where Red and NIR are spectral reflectance measurements corresponding to red and near infrared (NIR) parts of the spectrum; see, e.g., [1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 13, 14].

Related question. Assuming that the amount of life green vegetation can be uniquely determined by the values Red and NDVI, why this particular combination of these two values is the most adequate?

What we do in this paper. In this paper, we provide a theoretical explanation for the formula (1).

2 Towards an Explanation: General Analysis of the Problem

Different scales. When we process data, we process numerical values of the physical quantities. These numerical values depend not only on the quantity itself, they depends also on what measuring unit we select, and what starting point we select for measurement.

For example, to measure temperature, we can use Celsius (C) and Fahrenheit (F) scales. These scales use different measuring units – a 1-degree difference in the C scale corresponds to 1.8 degree difference in the F scale. These scales also use different starting points: 0 degrees on a C scale correspond to 32 degrees on the F scale.

In general, if we replace the original measuring unit by a new unit which is a times smaller, then all numerical values x get multiplied by a: $x \mapsto a \cdot x$. For example, if we replace meters by centimeters – a 100 times smaller unit – then, e.g., the original height of 1.7 m becomes $1.7 \cdot 100 = 170$ cm. Similarly, if we replace the original starting point with a new point which is b degrees smaller, then to all numerical values we add this value b: $x \mapsto x + b$. Thus, in general, if we change both the measuring unit and the starting point, we get a linear transformation $x \mapsto a \cdot x + b$.

Nonlinear rescalings. In many practical situations, we can also have nonlinear rescalings. For example, instead of describing the electric properties of an object, we can use resistance R or we can use conductivity $1/R$. Which of such nonlinear transformations are natural?

In general, linear transformations are natural. If we have a natural transformation from scale A to scale B, then the inverse transformation – from scale B to scale A – is also natural. If we have a natural transformation from scale A to scale B and another natural transformation from scale B to scale C, then their composition is a natural transformation from scale A to scale C. Thus, the class of all natural transformations is closed under composition and under taking an inverse, i.e., in mathematical terms, this means that this class should be a *transformation group*.

At any given moment of time, we can store and process only finitely many values. Thus, to effectively deal with different natural transformations, we should require that a natural transformation should be uniquely determined by the values of finitely

many parameters. In mathematical terms, this means that the class of all natural transformations should be *finite-dimensional*.

It is known – see, e.g., [9] and references therein – that for transformations from real numbers to real numbers, each finite-dimensional transformation group containing all linear transformations contains only fractional linear transformations. In other words, all natural transformations should be fractional linear, i.e., have the form

$$x \mapsto \frac{a \cdot x + b}{c \cdot x + d}, \tag{2}$$

for some constants a, b, c, and d.

Natural functions of two variables. For a function $z = f(x,y)$ of two variables:

- if we fix x, then we get a transformation $y \mapsto f(x,y)$, and
- if we fix y, then we get a transformation $x \mapsto f(x,y)$.

It is reasonable to say that a function $f(x,y)$ is natural if all these transformations are natural – and thus, fractional linear. Let us characterize all such functions.

3 First Result: Characterizing All Natural Functions of Two Variables

Reminder. Let us recall that a function $f(x,y)$ is called *bilinear* if it has the form

$$f(x,y) = a_0 + a_1 \cdot x + a_2 \cdot y + a_{12} \cdot x_1 \cdot x_2. \tag{3}$$

In these terms, we have the following result.

Definition 1. *We say that a function $f(x,y)$ is* natural *if the following two conditions are satisfied:*

- *for every x, the mapping $y \mapsto f(x,y)$ is fractional linear, and*
- *for every y, the mapping $x \mapsto f(x,y)$ is fractional linear.*

Proposition 1. *A function $f(x,y)$ is natural if and only if it is a ratio of two bilinear functions.*

Proof. If we fix the value of one of the variables, then a bilinear function becomes linear, and thus, the ratio of two bilinear functions becomes fractional linear. Thus, a ratio of two bilinear functions is natural in the sense of Definition 1.

So, to prove the proposition, it is sufficient to prove that every natural transformation is a ration of two bilinear functions. Indeed, naturalness means, in particular, that for every y, there exists values $a(y)$, $b(y)$, $c(y)$, and $d(y)$ for which

$$f(x,y) = \frac{a(y) \cdot x + b(y)}{c(y) \cdot x + d(y)}. \tag{4}$$

In the generic case, $d(y) \neq 0$, so we can delete both numerator and denominator by $d(y)$ and get a simpler expression

$$f(x,y) = \frac{A(y) \cdot x + B(y)}{C(y) \cdot x + 1}, \tag{5}$$

where we denoted

$$A(y) \stackrel{\text{def}}{=} \frac{a(y)}{d(y)}, \quad B(y) \stackrel{\text{def}}{=} \frac{b(y)}{d(y)}, \text{ and } C(y) \stackrel{\text{def}}{=} \frac{c(y)}{d(y)}.$$

(The case when $d(y) \equiv 0$ can be treated similarly.)

Since the function $f(x,y)$ is natural, for each x, the expression (5) is a fractional linear function of y. In particular, if we select three generic different values x_1, x_2, and x_3, we conclude that

$$f_i(y) = \frac{A(y) \cdot x_i + B(y)}{C(y) \cdot x_i + 1}, \tag{6}$$

where $f_i(y) \stackrel{\text{def}}{=} f(x_i, y)$ is the corresponding fractional linear function. Multiplying both sides of the formula (6) by the denominator, we conclude that

$$C(y) \cdot x_i \cdot f_i(x) + f_i(y) = A(y) \cdot x_i + B(y), \tag{7}$$

i.e., equivalently, that

$$x_i \cdot A(y) + B(y) - f_i(y) \cdot x_i \cdot C(y) = -f_i(y). \tag{8}$$

So, for each y, we have the system of three linear equations to determine the three unknowns $A(y)$, $B(y)$, and $C(y)$. According to Cramer's rule, the solution to a system of linear equations is a ratio of two determinants – which are polynomials in terms of the coefficients. In mathematics, ratios of polynomials are called rational functions. So, the solution to a system of linear equations is a rational function of all the coefficients. In our case, x_i, 1, and -1 are constants, and $f_i(y) \cdot x_i$ is a rational function. So, each of the coefficients $A(y)$, $B(y)$, and $C(y)$ is a rational function of a rational function – and thus, a rational function itself. In other words:

$$A(y) = \frac{P_A(y)}{Q_A(y)}, \quad B(y) = \frac{P_B(y)}{Q_B(y)}, \quad C(y) = \frac{P_C(y)}{Q_C(y)} \tag{9}$$

for some polynomials $P_A(y)$, $Q_A(y)$, $P_B(y)$, $Q_B(y)$, $P_C(y)$, and $Q_C(y)$. Substituting the expressions (9) into the formula (5), we conclude that

$$f(x,y) = \frac{\dfrac{P_A(y)}{Q_A(y)} \cdot x + \dfrac{P_B(y)}{Q_B(y)}}{\dfrac{P_C(y)}{Q_C(y)} \cdot x + 1}. \tag{10}$$

By using the usual formulas for addition and division of fractions, we conclude that

$$f(x,y) = \cfrac{\cfrac{P_a(y) \cdot Q_B(y) + x \cdot P_B(y) \cdot Q_A(y)}{Q_A(y) \cdot Q_B(y)}}{\cfrac{x \cdot P_C(y) + Q_C(y)}{Q_C(y)}} =$$

$$\frac{P_A(y) \cdot Q_B(y) \cdot Q_C(y) + x \cdot P_B(y) \cdot Q_A(y) \cdot Q_C(y)}{x \cdot Q_A(y) \cdot Q_B(y) \cdot P_C(y) + Q_A(y) \cdot Q_B(y) \cdot Q_C(y)}. \tag{11}$$

In other words, the function $f(x,y)$ is a ratio of two polynomials which are linear in x. These two polynomials may have a non-constant common divisor, which is either linear in x or does not depend on x at all. Dividing both numerator and denominator by the greatest common divisor of these two polynomials, we conclude that

$$f(x,y) = \frac{a_0(y) + x \cdot b_0(y)}{x \cdot c_0(y) + d_0(y)} \tag{12}$$

for some polynomials $a_0(y)$, $b_0(y)$, $c_0(y)$, and $d_0(y)$.

Similarly, from the fact that for each y, the mapping $x \mapsto f(x,y)$ is fractional linear, we conclude that

$$f(x,y) = \frac{A_0(x) + y \cdot B_0(x)}{y \cdot C_0(x) + D_0(y)} \tag{13}$$

for some polynomials $A_0(x)$, $B_0(x)$, $C_0(y)$, and $D_0(x)$. Equating the right-hand sides of the formulas (12) and (13), we conclude that

$$\frac{a_0(y) + x \cdot b_0(y)}{x \cdot c_0(y) + d_0(y)} = \frac{A_0(x) + y \cdot B_0(x)}{y \cdot C_0(x) + D_0(x)}. \tag{14}$$

Multiplying both sides of this equality by both denominator, we conclude that

$$(a_0(y) + x \cdot b_0(y)) \cdot (y \cdot C_0(x) + D_0(x)) = (x \cdot c_0(y) + d_0(y)) \cdot (A_0(x) + y \cdot B_0(x)). \tag{15}$$

The left-hand side of this equality is divisible by $a_0(y) + x \cdot b_0(y)$, which means that the right-hand side must be divisible by the same expression. By our construction of the expression (12), this sum has no common factors with

$$x \cdot c_0(y) + d_0(y).$$

Thus, it must divide $A_0(x) + y \cdot B_0(x)$.

Similarly, the right-hand side of the equality (15) is divisible by

$$A_0(x) + y \cdot B_0(x).$$

Since by our construction of the expression (13), this sum has no common factors with $y \cdot C_0(x) + D_0(x)$. Thus, it must divide $a_0(y) + x \cdot b_0(y)$.

So, the polynomials $a_0(y) + x \cdot b_0(y)$ and $A_0(x) + y \cdot B_0(x)$ must divide each other and thus, they should differ only by a multiplicative constant c:

$$a_0(y) + x \cdot b_0(y) = c \cdot A_0(x) + y \cdot c \cdot B_0(x) \tag{16}$$

for all x and y. In particular, for two different values x_1 and x_2, we conclude that

$$a_0(y) + x_i \cdot b_0(y) = c \cdot A_0(x_i) + y \cdot c \cdot B_0(x_i). \tag{17}$$

So, we have two linear equations with constant coefficients for determining two unknowns $a_0(y)$ and $b_0(y)$. Thus, the solution is a linear combination of the right-hand sides. Right-hand sides are linear functions of y, so both $a_0(y)$ and $b_0(y)$ are linear functions of y: $a_0(y) = a_{00} + a_{01} \cdot y$ and $b_0(y) = a_{10} + a_{11} \cdot y$, for some coefficients a_{ij}. Thus, the numerator of the expression (12) has the form

$$a_0(y) + x \cdot b_0(y) = a_{00} + a_{10} \cdot x + a_{01} \cdot y + a_{11} \cdot x \cdot y. \tag{18}$$

In other words, this numerator is a bilinear function.

Similarly, we can conclude that the denominator of the expression (12) is a bilinear function:

$$x \cdot c_0(y) + d_0(y) = c_{00} + c_{10} \cdot x + c_{01} \cdot y + c_{11} \cdot x \cdot y \tag{19}$$

for some coefficients c_{ij}. Thus, according to the formula (12), the function $f(x,y)$ is indeed equal to the ratio of two bilinear functions:

$$f(x,y) = \frac{a_{00} + a_{10} \cdot x + a_{01} \cdot y + a_{11} \cdot x \cdot y}{c_{00} + c_{10} \cdot x + c_{01} \cdot y + c_{11} \cdot x \cdot y}. \tag{20}$$

The proposition is proven.

Comment. A similar result can be similarly proven for a function of several variables. Let us recall that a function $f(x_1, \ldots, x_n)$ is called *multi-linear* if it has the form

$$f(x_1, \ldots, x_n) = \sum_{S \subseteq \{1,\ldots,n\}} a_S \cdot \prod_{i \in S} x_i. \tag{21}$$

In these terms, we have the following result.

Definition 2. *We say that a function $f(x_1, \ldots, x_n)$ is* natural *if for all i and for all possible values $x_1, \ldots, x_{i-1}, \ldots, x_{i+1}, \ldots, x_n$, the mapping*

$$x_i \mapsto f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$$

is fractional linear.

Proposition 2. *A function $f(x_1, \ldots, x_n)$ is natural if and only if it is a ratio of two multi-linear functions.*

4 Scale Invariance

Requirement. The values Red and NDVI depend on the solar angle. For a different angle, the same situation leads to the values Red and NDVI both multiplied by the same constant. Thus, to make a characteristic that does not depend on the solar angle, we want to select a characteristic $f(\text{NIR}, \text{Red})$ that should not change if we simply change the angle, i.e., for which

$$f(c \cdot x, c \cdot y) = f(x, y)$$

for all possible values of c, x, and y. Such functions are known as *scale-invariant.*

The following result characterizes all scale-invariant natural functions.

Proposition 3. *A natural function $f(x,y)$ which is not identically constant is scale-invariant if and only if it has the form*

$$f(x,y) = \frac{a_{10} \cdot x + a_{01} \cdot y}{c_{10} \cdot x + c_{01} \cdot y}. \tag{22}$$

Proof. One can easily prove that each expression (22) is scale-invariant. Vice versa, let us assume that a natural function is scale-invariant. We know, from Proposition 1, that all natural functions are described by expression (20). Thus, scale invariance means that for all x and y, we have

$$f(x,y) = \frac{a_{00} + a_{10} \cdot x + a_{01} \cdot y + a_{11} \cdot x \cdot y}{c_{00} + c_{10} \cdot x + c_{01} \cdot y + c_{11} \cdot x \cdot y} =$$

$$\frac{a_{00} + c \cdot a_{10} \cdot x + c \cdot a_{01} \cdot y + c^2 \cdot a_{11} \cdot x \cdot y}{c_{00} + c \cdot c_{10} \cdot x + c \cdot c_{01} \cdot y + c^2 \cdot c_{11} \cdot x \cdot y}. \tag{23}$$

To prove the proposition, we need to prove that $a_{00} = c_{00} = 0$ and that $a_{11} = c_{11} = 0$.

Let us prove that $a_{00} = c_{00} = 0$. Indeed, if at least of these two coefficients is different from 0, then for $c \to 0$, the right-hand side of the equality (23) becomes a constant a_{00}/c_{00}, so we conclude that the function $f(x,y)$ is constant.

Similarly, if at least one of the coefficients a_{11} and c_{11} is different from 0, then in the limit $c \to \infty$, the right-hand side of the equality (23) becomes a constant a_{00}/c_{00}, so we conclude that the function $f(x,y)$ is constant.

So, if the function $f(x,y)$ is not constant, all these four coefficients should be equal to 0. Thus, Proposition 3 is proven.

Discussion. We have almost got the desired expression (1). To get even closer to the expression (1), let us take into account that the NDVI index changes from -1 to 1. If we require that the range of the function is exactly $[-1, 1]$, then we get the following result.

Proposition 4. *A natural scale-invariant function $f(x,y)$ whose range of values for $x, y \geq 0$ is $[-1, 1]$ has the form*

$$f(x,y) = \frac{x - c \cdot y}{x + c \cdot y} \text{ or } f(x,y) = -\frac{x - c \cdot x}{x + c \cdot y}. \tag{24}$$

Proof. If we divide both the numerator and the denominator of the expression (22) by x, we get

$$f(x,y) = F(z) \stackrel{\text{def}}{=} \frac{a_{10} + a_{01} \cdot z}{c_{10} + c_{11} \cdot z}, \text{ where } a = \frac{y}{x}. \tag{25}$$

The fact that the range of this function is between -1 and 1 means that this function is defined for all z. In this case, the fractional linear function (25) is monotonic, so its range when $z \in [0, \infty]$ is simply an interval bounded by the values of this function at the endpoints $z = 0$ and $z = \infty$. So, the fact that the range is equal to $[-1, 1]$ means that one of the values $F(0)$ and $F(\infty)$ should be equal to 1, and another value to -1.

If $F(0) = -1$, then for its opposite $G(z) \stackrel{\text{def}}{=} -F(z) = -f(x,y)$, we have $G(0) = -F(0) = 1$. So, without losing generality, we can consider the case when $F(0) = 1$. Substituting the value $z = 0$ into the right-hand side of the formula (25), we conclude that $a_{10}/c_{10} = 1$, i.e., that $a_{10} = c_{10}$. Dividing both the numerator and the denominator of the expression (25) by this common value $a_{10} = c_{10}$, we conclude that

$$F(z) = \frac{1 + a \cdot z}{1 + c \cdot z}, \text{ where } a = \frac{a_{01}}{a_{10}} \text{ and } c = \frac{c_{01}}{c_{10}}. \tag{26}$$

This ratio has to be defined for all z, so we must have $c \geq 0$ – otherwise, if $c < 0$, this expression would not be defined for $z = -1/c$.

For $z = \infty$, we have $F(\infty) = -1$, so $a/c = -1$, thus $a = -c$. The proposition is proven.

Conclusion. We (almost) explained why NDVI is a relevant characteristics.

Comment. Similar argument can – partly – explain the effectiveness of fractional-linear similarity coefficients like Jaccard index

$$s(A,B) = \frac{\mu(A \cap B)}{\mu(A) + \mu(B) - \mu(A \cap B)},$$

where $\mu(S)$ denotes the measure of the set S; see, e.g., [15].

Acknowledgments

References

1. R. E. Crippen, "Calculating the vegetation index faster", *Remote Sensing of Environment*, 1990, Vol. 34, pp. 71–73.
2. D. M. Gates, *Biophysical Ecology*, Springer-Verlag, New York, 1980.
3. B. N. Holben, "Characteristics of maximum-value composite images from temporal AVHRR data", *International Journal of Remote Sensing*, 1986, Vol. 7, No. 11, pp. 1417–1434.
4. A. R. Huete, "A soil-adjusted vegetation index (SAVI)", *Remote Sensing of Environment*, 1988, Vol. 25, pp. 53–70.
5. Y. J. Kaufman and D. Tanre, "Atmospherically resistant vegetation index (ARVI) for EOS-MODIS", *Proceedings of the IEEE International Geosciences and Remote Sensing Symposium'92*, New York, 1992, pp. 261–270.
6. F. J. Kriegler, W. A. Malila, R. F. Nalepka, and W. Richardson, "Preprocessing transformations and their effects on multispectral recognition", *Proceedings of the Sixth International Symposium on Remote Sensing of Environment*, 1969, pp. 97–131.
7. S. M. Labib, S. Lindley, and J. J. Huck, "Scale effects in remotely sensed greenspace metrics and how to mitigate them for environmental health exposure assessment", *Computers, Environment and Urban Systems*, 2020, Vol. 82, Paper 101501, doi:10.1016/j.compenvurbsys.2020.101501.
8. R. B. Myneni, F. G. Hall, P. J. Sellers, and A. L. Marshak, "The interpretation of spectral vegetation indexes", *IEEE Transactions on Geoscience and Remote Sensing*, 1995, Vol. 33, pp. 481–486.
9. H. T. Nguyen and V. Kreinovich, *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht, 1997.
10. H. Nouri, S. Anderson, P. Sutton, S. Beecham, P. Nagler, C. J. Jarchow, and D. A. Roberts, "NDVI, scale invariance and the modifiable areal unit problem: An assessment of vegetation in the Adelaide Parklands", *Science of the Total Environment*, 2017, Vol. 584–585, pp. 11–18. doi:10.1016/j.scitotenv.2017.01.130.
11. B. Pinty and M. M. Verstraete, "GEMI: a non-linear index to monitor global vegetation from satellites", *Vegetatio*, 1992, Vol. 101, No. 1, pp. 15–20, doi:10.1007/bf00031911.
12. A. J. Richardson and C. L. Wiegand, "Distinguishing vegetation from soil background information", *Photogrammetric Engineering and Remote Sensing*, 1977, Vol. 43, pp. 1541–1552.
13. J. W. Rouse, R. H. Haas, J. A. Scheel, and D. W. Deering, "Monitoring vegetation systems in the Great Plains with ERTS", *Proceedings of the 3rd Earth Resource Technology Satellite (ERTS) Symposium*, 1974, Vol. 1, pp. 48–62.
14. P. J. Sellers, "Canopy reflectance, photosynthesis, and transpiration", *International Journal of Remote Sensing*, 1985, Vol. 6, pp. 1335–1372.
15. P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to Data Mining*, Pearson, New York, 2018.