Key Biodiversity Areas

Details on the Calculation of the different Scores

29/01/2025

© Eviden SAS

Overall Condition Score (OCS)

Overview

The Overall Condition Score is used to present complex information from multiple indicators as a single metric that measures progress toward a goal and to facilitate decision-making. This score is made up of different sub-scores and tries to be a score that contains the information content of those different pieces of information in one.

These different sub-scores are:

- **Infrastructure Score**

  This score indicates how likely an infrastructure object is to exist in a specific geographical area. This score is calculated from the object recognition and how likely it is that the AI recognized one or more objects (boma, fence, building, etc.) as such.

- **NDVI Outlier Score**

  The NDVI (Normalized Difference Vegetation Index) outlier value is a value that compares the current NDVI value with the average value from previous years for the same month.). Large deviation will be called an outlier.

- Normalized Burn Ratio (NBR) – will be added in the future

  The Normalized Burn Ratio is used to identify burned areas and provide a measure of burn severity and is calculated as a ratio between the NIR (near-infra red) and SWIR(short-wave infra red) values.

© Eviden SAS

Overall Condition Score (OCS)

Requirements and Limitation

The calculation of an Overall Condition Score includes requirements that should/must be adhered to in order to be accepted as an acceptable and trustworthy score.

Following you can find all requirements for an OCS score:

- Req_1: The OCS is a compound index which will be calculated by the usage of sub-indices which are thematically group variables
- Req_2: Only the sub-scores (range of [0:1]) are permitted for the calculation: Infrastructure_Score, NDVI_OutlierScore
- Req_3: Weighting the sub-indices can be done (if so, then the sum of all weights must be 100% or 1)
- Req_4: OCS is a value in the range of 0 to 1
- Req_5: If all sub-scores are 0, the OCS must have a value of 0
- Req_6: If all sub-scores are 1, the OCS must have a value of 1
- Req_7: High values of one sub-score must not be canceled out by low values of another sub-score

© Eviden SAS

Overall Condition Score

Calculation

After testing different calculation methods by reviewing the results of the different methods with varying input sub-score values and checking which method fulfills the most requirements, the following method was chosen:

Max-Adjusted Geometric Mean

It combines the maximum value with a weighted geometric mean to ensure fairness and compliance with all requirements.

$$\text{OCS} = Max(v_1, \dots, v_n) + \sqrt[n]{v_1 \times \cdots \times v_n} \times (1 - Max(v_1, \dots, v_n))$$

Notation:

– OCS is the calculated Overall Condition Score
– v is the input vector (sub-score) and v is in the range of [0:1]
– n is the number of different sub-scores (currently n = 2)

This formular consists out of three parts:

- **Maximum value** ($Max(v_1, \dots, v_n)$)
  This ensures the OCS is at least equal to the maximum of all input variables
- **Geometric mean** ($\sqrt[n]{v_1 \times \cdots \times v_n}$)
  Combines all variables in a balanced way, avoiding dominance by a single variable. It is also preferred when combining independent variables as it respects their scale and avoids overemphasizing either extremely high or low values.
- **Weight adjustment** ($1 - Max(v_1, \dots, v_n)$)
  scales the geometric mean's contribution based on the range left after accounting for the maximum value

© Eviden SAS

Infrastructure Score

Overview

This score is used to measure the human made infrastructure in each area by converting the number of objects and their confidence score in an area into a score between 0 and 1. Zero means there is no detectable human-made infrastructure, and one means there is some kind of hotspot where the human footprint is highly visible. To detect changes, monthly object detections are done to monitor the key biodiversity areas.

The object detection is done by an AI model which was trained with labelled images from the Dakatcha Woodland and the Amboseli National Park. The images are provided monthly by Planet.com and have a resolution of approximately 3m/pixel.

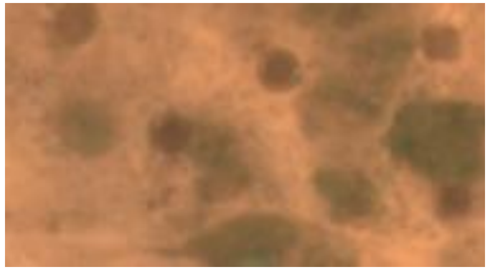
The classes of objects that where labelled and that the model was trained on are **bomas**, **buildings** and **fields**. The object class fence will be added in the future.

© Eviden SAS

Infrastructure Score

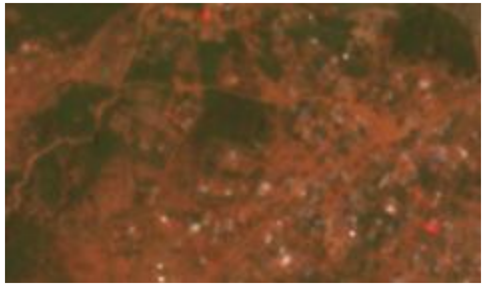
Object Classes

Boma



- The image above shows a cluster of Bomas in the Amboseli National Park (the dark circles)
- A boma is a livestock enclosures
- Very frequent in the Amboseli National Park, uncommon in the Dakatcha Woodland

Building



- The image above shows a small town in Amboseli
- The coloured dots are the individual buildings
- These three images are the actual satellite images from Planet.com which were included in the training

Field

- The image above shows fields in the Dakatcha Woodland
- Field stands for any human-made clearings that are mainly used for agricultural purposes
- Fields shape the landscape of the Dakatcha Woodland

Infrastructure Score

Object Score Calculation

For each hexagon and each object class the number of detected objects and the object class score is saved. The object class score is a metric which combines the number of objects in that area with the corresponding confidence scores and stays in the range [0:1]. After comparing different methods, the following formular was chosen:

$$x = Min\{1, \mu \cdot \ln(1 + n)\}$$

Notation:
- x is the calculated object class score
- n is the saved object count
- μ is the average confidence, so the sum of all the confidence values of the detected objects divided by the count
- ln() is the natural log

This calculation is sufficient for object classes where the number of detected objects is critical, like for buildings and bomas, however, for fields the area is deemed to be more crucial. Therefore, fields are calculated differently with the following formular that depends only on the area of the field.

$$x = Min\{1, \ln(f \cdot \pi + 1)\}$$

Notation:
- x is the calculated object class score
- f is the field factor that depends on the hexagon size (= zoom level)
- π is the ratio of the intersection area of the field with the hexagon divided by the hexagon area

© Eviden SAS

Infrastructure Score

Object Score Calculation – Zoom Levels

To minimize loading times, there are three different zoom levels of the hexagons. So, the more one zooms out, the bigger the hexagons become. The biggest hexagons have a circumradius of 10000m, the medium ones of 1000m and the smallest ones of 100m. Even though the bigger hexagons are only there to give a gross estimation of where the hot spots are, to then zoom in for more information, the object score calculation needs to be adapted to these zoom levels, because it would otherwise be completely meaningless. As the current calculation for buildings and bomas depends on the number of detected objects, bigger hexagons would almost always have a score of 1 with the same calculation. On the other hand, the field score would almost always be very close to 0, because of the huge hexagon area. The following adaptations are implemented for each zoom level:

Hexagon Circumr.	Boma and Building score calculation $x = Min\{1, \mu \cdot \ln(1+n)\}$	Field score calculation $x = Min\{1, \ln(f \cdot \pi + 1)\}$
100	$x = Min\{1, \mu \cdot \ln(1+n)\}$	$x = Min\{1, \ln(2 \cdot \pi + 1)\}$ x gets 1 if the ratio is ~ 0.85 or higher
1000	$x = Min\{1, \mu \cdot \log_b(1+n)\}$ with the base b = e + 3	$x = Min\{1, \ln(8 \cdot \pi + 1)\}$ x gets 1 if the ratio is ~ 0.21 or higher
1000	$x = Min\{1, \mu \cdot \log_b(1+n)\}$ with the base b = e + 33	$x = Min\{1, \ln(128 \cdot \pi + 1)\}$ x gets 1 if the ratio is ~ 0.013 or higher

© Eviden SAS

Infrastructure Score

Infrastructure Score Calculation

The overall infrastructure score is calculated by combining all the object class scores in the following way:

$$\text{Infrastructure\_score} = 1 - ((1 - \text{boma\_score}) * (1 - \text{building\_score}) * (1 - \text{field\_score}))$$

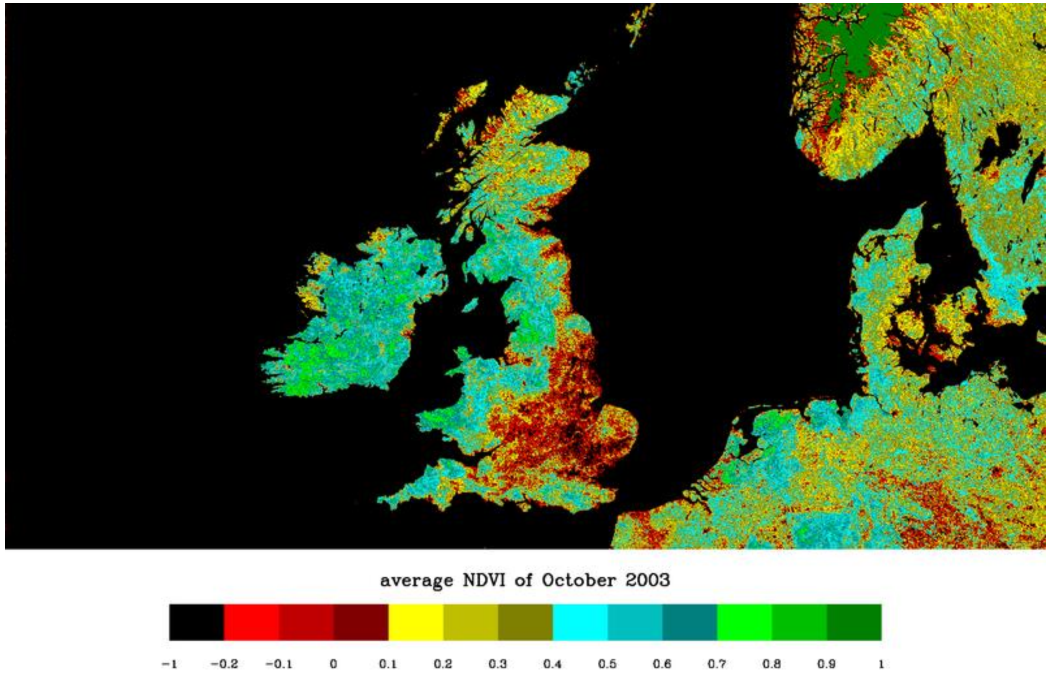
As for now, the **object classes are not weighted** differently and are calculated into the infrastructure score equally. Including the weight in the calculation obviously has an impact on the result, because the weighting reflects a prioritization of the object classes. Weights can be added in the future by experts who know the domain and believe a prioritization is beneficial.

© Eviden SAS

Normalized Difference Vegetation Index (NDVI)

Overview

The Normalized Difference Vegetation Index (NDVI) is a widely used metric for quantifying the health and density of vegetation using sensor data. It's calculated from spectrometric data at two specific bands: red and near-infrared. Essentially, NDVI helps assess vegetation cover by normalizing green leaf scattering in the near-infrared wavelength and chlorophyll absorption in the red wavelength.



https://commons.wikimedia.org/wiki/File:NDVI_102003.png

EVIDEN Atos
© Eviden SAS

NDVI

Calculation

There are 3 different NDVI values and the NDVI outlier score that are displayed. The heatmap displays the NDVI outlier score, which shows the deviation of the value from usual/expected NDVI value of this month. When one hexagon is then selected, the expected and the measured NDVI value of this area can be seen.

- **Measured NDVI value**
  The NDVI calculation is performed for each pixel within the image. Calculation of NDVI will be done through subtraction and division of the values.

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Notation:
- $NDVI$ is the calculated NDVI value for one pixel
- NIR is the pixel value of band 4 of a TIF-file: Containing Near Infra-red (NIR) data
- RED is the pixel value of band 3 of a TIF-file: Containing RGB data

The result is an array of NDVI values within a range of -1 and +1. Values closer to +1 indicate a greater density of vegetation of higher level of “greenness”, values less than 0 are, by definition, water or non-vegetated features and are not used for further calculation. The valid NDVI values for each hexagon are now averaged and the result is the measured NDVI value for that hexagon.

© Eviden SAS

NDVI

Calculation

- **Expected NDVI value**
  This is the mean value of all NDVI values from previous years of the same/current month.

$$ndvi_{expected} = \frac{\sum_i^n x_i}{n}$$

Notation:
- n is the number of years where the NDVI of this month was already measured and can be added to the historical data
- x_i is the i-th NDVI value of all n historic NDVI values

- **NDVI variance**
  This value is the variance of the expected NDVI values measured in the past. Hence, it is the mean variation of the historical values of the same month.

$$ndvi_{var} = \frac{\sum_i^n (x_i - ndvi_{expected})^2}{n}$$

Notation:
- n is the number of previous years of the current month
- x_i is the i-th NDVI value of all n historic NDVI values

© Eviden SAS

NDVI

Calculation

- **NDVI outlier score**
  This is a scoring value which is calculated by the current value in comparison to the previous year's values for the current month. A strong deviation from the historical values is then referred to as an outlier. An outlier is an observation that lies abnormally far away from other values in a dataset. The output range is in the interval from -1 to +1. Zero means that there is no outlier, and the actual value is in the range of the historical values. A negative value means that the outlier is smaller(browner) than the measured value, otherwise larger (greener). The current implementation is based on the calculation of the z-score.
- The **z-score** can be used to identify outliers by measuring how many standard deviations a single observation or data point is away from the mean of the distribution.
  Example: A z-score of ±2: Values within two standard deviations from the mean are less common and comprise about 95% of the data.
  The result of the z-score calculation is scaled by a continuous function (sigmoid) and resulting in values between 0 (no outlier) and 1(certain outlier).

$$\text{zScore} = \frac{ndvi_{measured} - ndvi_{expected}}{\sqrt{ndvi_{var}}}$$

$$ndvi_{os} = \frac{\text{sgn}(zScore)}{1 + e^{-k(|zScore| - z0)}}$$

Notation:

- z_0 threshold
  if zScore = z_0 then the outlier score is 0.5 (middle of the interval)
- k is the slope of the sigmoid function
  The larger the value the faster get result is getting to the limits of [0,1]
- Suggestion for the parameters: k=3 and z0=1.5
- sgn(...) returns the value -1 if less than 0, otherwise +1

© Eviden SAS

NDVI

Calculation

- **NDVI OS calculation: The early years**
  The calculations described above represent the general case. Here, special attention is now paid to the first five years. The reason for this is that the calculation of the values (OS, mean values, variance, zScore) is initially considered separately and there are various ways of dealing with this. The following method, which is also implemented, takes the first years into account. It is assumed that 5 years (2020-2024) are already available when the data is created. Therefore, all five NDVI measurements for this entire period are used to calculate the mean and variance and are not changed. From 2025 onwards, all past measurements will be included in the calculation and the measurements will no longer be treated specially.
- **Test data**
  The test data has a historic testing period (starting from 2020 – 2024).
  For this period, the mean value of the measured data is 0.676 and the variance is 0.0091

V1: Test without special treatment

```
V1: year=2020 measured=0.7700  mean=0.7700  var=0.0000  OS=0.000
V1: year=2021 measured=0.6000  mean=0.7700  var=0.0010  OS=1.000
V1: year=2022 measured=0.6000  mean=0.6850  var=0.0145  OS=0.040
V1: year=2023 measured=0.7900  mean=0.6567  var=0.0096  OS=0.362
V1: year=2024 measured=0.6200  mean=0.6900  var=0.0109  OS=0.035
V1: year=2025 measured=0.7700  mean=0.6760  var=0.0091  OS=0.113
V1: year=2026 measured=0.6100  mean=0.6917  var=0.0088  OS=0.075
V1: year=2027 measured=0.7100  mean=0.6800  var=0.0083  OS=0.009
```

V2: Test special treatment (implemented version)

```
V2: year=2020 measured=0.7700  mean=0.6760  var=0.0091  OS=0.113
V2: year=2021 measured=0.6000  mean=0.6760  var=0.0091  OS=0.056
V2: year=2022 measured=0.6000  mean=0.6760  var=0.0091  OS=0.056
V2: year=2023 measured=0.7900  mean=0.6760  var=0.0091  OS=0.227
V2: year=2024 measured=0.6200  mean=0.6760  var=0.0091  OS=0.025
V2: year=2025 measured=0.7700  mean=0.6760  var=0.0091  OS=0.113
V2: year=2026 measured=0.6100  mean=0.6917  var=0.0088  OS=0.075
V2: year=2027 measured=0.7100  mean=0.6800  var=0.0083  OS=0.009
```

© Eviden SAS