Practical guide

BICYCLE COUNTING: HOW TO CLEAN UP THE DATA?















INTRODUCTION

Why checking the quality of counting data and cleaning them is important? For two reasons: to identify breakdowns and possible equipment maintenance needs and to increase the reliability of the analyses.

Without looking at the data in detail and in the absence of periodic manual count backtracking, it is sometimes difficult to detect problems with counter operation. Close and regular monitoring of the data allows the identification of the main failures of the counters.

As with any data analysis, it is imperative to check the quality of the counting data and to clean it. **This avoids the inclusion of outliers in the analyses and the drawing of inappropriate conclusions**. For example, a broken counter that produces data at 0 for one month. By comparing the raw data with the previous year, it can be concluded that the number of passages at this location is decreasing, while it may be increasing.

Data cleansing: what is an outlier, how to identify and deal with it?

The cleaning of the counting data starts with the detection of anomalies. But what is an anomaly for bicycle count data? We will first answer this question, then present the different types of anomalies that can be encountered, how to detect them and what actions to take afterwards.

1. What is a counting anomaly?

An anomaly is defined as any data that deviates from the trend usually observed. Anomaly does not necessarily mean "bad data".

What may at first glance be considered an anomaly may turn out to be completely justified (e.g. a peak in attendance linked to a particular event).

2. What are the main types of anomalies?

There are 6 main types of outliers:

- Missing data
- Data at 0
- The peaks or troughs
- Increases or decreases
- Inconsistencies in the time profile
- Inconsistencies in the distribution of data by counting direction

3. How to detect counting anomalies?

To detect counting anomalies, it is important to analyse the data in the most detailed manner possible, even if only a global analysis is produced. Thus, it is advisable to have a look at the data:

- . per hour
- . per traffic direction when available

When outliers are detected, it is also advisable to check the data for the previous and following days more closely. Sometimes these may appear/disappear gradually. It will then be important to treat these periods in the same way as the heart of the anomaly.

CONTENTS

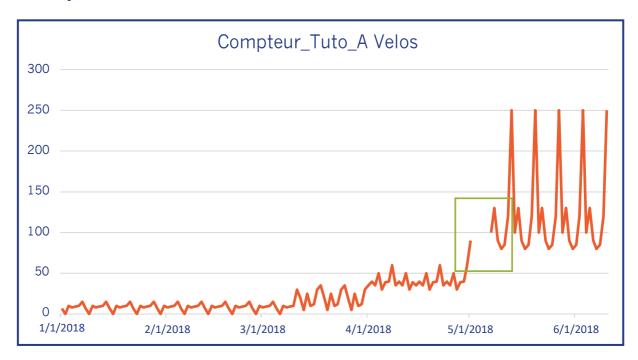
1.	MISSING DATA	4
l.	Characteristics and visualization	4
.	Potential technical or field explanations	5
III.	Data processing	6
2.	LES DONNÉES À ZÉRO	7
l.	Characteristics and visualization.	7
II.	Potential technical or field explanations	8
III.	Data processing	9
3.	LES PICS OU LES CREUX	. 10
l.	Characteristics and visualization	10
II.	Potential technical or field explanations	13
III.	Data processing	13
4.	LES HAUSSES OU BAISSES	. 14
l.	Characteristics and visualization.	14
II.	Potential technical or field explanations	17
III.	Data processing	18
5.	LES INCOHÉRENCES DANS LE PROFIL HORAIRE	. 19
l.	Characteristics and visualization.	19
.	Potential technical or field explanations	20
III.	Data processing	20
6.	LES INCOHÉRENCES DANS LA RÉPARTITION DES DONNÉES PAR SENS DE COMPTAGE.	. 22
l.	Characteristics and visualization	. 22
Ш	Detential technical or field avalenations	2/
	Potential technical or field explanations	24

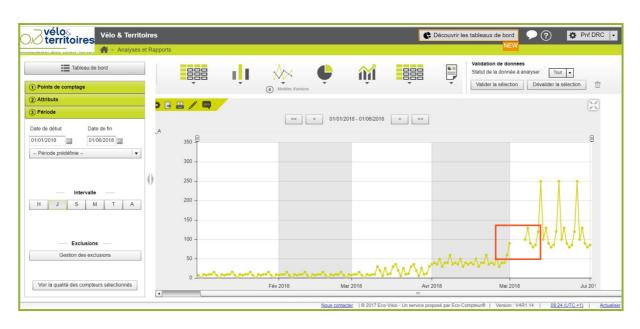


MISSING DATA

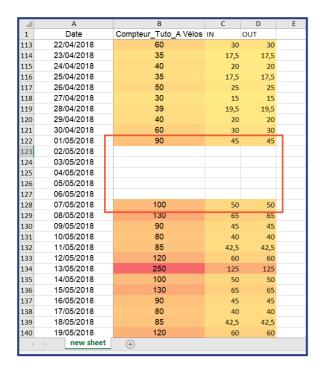
I. Characteristics and visualization

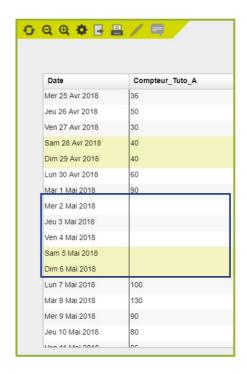
Missing data is characterised by an empty area (a hole) in the data set. It can be temporary (from one hour to several weeks) with or without data recovery. In graphical form, the absence of data takes the form of an interruption in the curve, with or without recovery.





In table form (in Eco-Visio or Excel), no data will be available for the period concerned. In Excel, these periods without data are easy to view by using a conditional formatting by colour shade. The period with missing data will not be coloured, while the others will be.





II. Potential technical or field explanations

For 'Eco-Counter' devices equipped with GSM data transmission module:

Missing data from one hour to several days with recovery:

Network coverage or a one-time network problem (related to the SIM card provider) did not allow transmission during these days. It is necessary to restart transmission for the missing days. To do this, technical support must be called within a maximum of 80 days for Eco GSM and COMBO 1 counters and 100 days for COMBO 2 counters. Beyond that, it will be necessary to go to the counting site to manually retrieve the data.

Missing data without recovery:

This may be due to a battery problem, for example. You can check the battery charge level on Eco-Visio.

- If the level is low, it is time to change the battery pack of your GSM module and to manually retrieve the missing data. The data is stored in the counter for a maximum of 11 months to 88 months depending on the counter model.
- If the indicated charge level is correct, it is necessary to use the support to remotely
 check the correct operation of the counter. If the remote intervention does not allow
 diagnosis, it will be necessary to go to the counting site to check the operation of the
 counter.

For counters connected to circulation management systems or urban power supply cabinets: temporary missing data (a few hours) can sometimes be linked to maintenance operations or network outages on the hosting structures.

In any case, in addition to the hardware failure, missing data may also be due to track work that would have damaged the counter. Ask the relevant road engineering departments to check the work schedule.

III. Data processing

If you notice missing data over a period of time, you can follow these steps:

- 1. Contact Eco-Counter technical support to understand and solve the problem. On a counter equipped with GSM that has not emitted, Eco-Compteur will be able to recover a data history of up to 100 days.
- **2. Read the data on your counter** directly on site. Depending on the configuration of the counter, the system stores between 11 months and more than 3 years of data.
- 3. If a technical failure is found and the counter has not recorded any data, a data reconstruction can be considered:
- For a few hours of missing data: you can use the automatic reconstruction feature available in Eco-Visio. This calculates for each hour the average of the values of the previous four weeks.
- For several days of missing data: a reconstruction can be carried out using counters with a high correlation of data (usually nearby counters). This reconstruction can be carried out by you via an import in Excel format, or on request to Eco-Compteur (on quote).

1 M

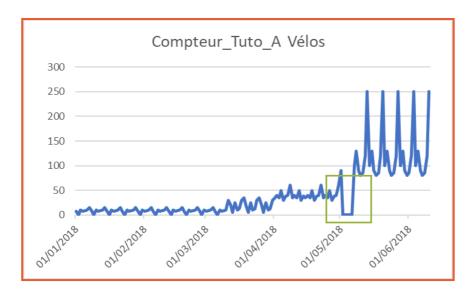
DATA AT ZERO

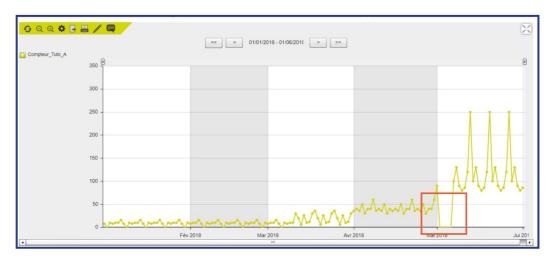
I. Characteristics and visualization

As shown in the two graphs below, null data is characterized by a data sequence of 0 in the dataset. This sequence can be temporary (from a few hours to several days or weeks) with or without data recovery.

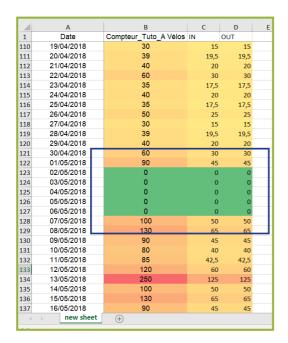
Data at 0 are not considered as outliers when they last a few hours during normal times (e.g. at night) or one or more days during normal periods (e.g. during flood periods for routes along rivers or in winter in mountainous or rural areas). On the other hand, data at 0 can be outliers when they are in the middle of the day when there is usually a lot of traffic.

In graphical form, the data at 0 will take the form of a curve that merges with the abscissa axis (horizontal), with or without repetition.





In the form of a table (in Eco-Visio or Excel), there will simply be a sequence of 0's that can be viewed for the period in question. In Excel, these periods of 0 are easily visualised using a conditional formatting by colour shade. They will appear in green while the other data will be yellow, orange or red. With Eco-Visio, the identification requires more vigilance (they can be identified by sorting the data in ascending order).





II. Potential technical or field explanations

Technical:

- A period at 0 without a return to a normal situation may indicate a battery failure, a hardware problem.
- A period at 0 with a return to a normal situation may indicate that the battery has been faulty and changed by the counter owner, or that a problem with the hardware has been resolved.

Field:

- A period of 0 may occur during work on the route. The route is completely cut off and no more cyclists pass over the counting site.
- A period at 0 can also occur when the route is cut off for natural reasons (flooding, snow) or because the weather conditions are not favourable for the cycling pratice.
- A period of 0 will also be visible if the ZELT loops have been sawn off or torn off during track work. This can also occur when the roadway has been covered with asphalt which prevents the sensor from detecting bicycles.

In any case, in addition to hardware failure, a data sequence at zero may be due to a break in the route, for example, for work. Ask the technical services of the relevant road authorities to check the work schedule.

III. Data processing

If you set a data period to zero, in case of technical problems on the counter, you can follow these steps:

- 1. Contact Eco-Counter technical support to understand and solve the problem.
- 2. If a technical failure is proven, a data reconstruction can be considered:
- For a few hours: you can use the automatic reconstruction feature available in Eco-Visio. This calculates for each hour the average of the values of the previous four weeks.
- For several days of data: a reconstruction can be carried out using counters with a high correlation of data (usually nearby counters). This reconstruction can be carried out by you via an import in Excel format, or on request to Eco-Compteur (on quote).

In the event of a route interruption, the data at zero corresponds to a real event. If there was no deviation proposed with a counter placed on it. It is preferable not to fill the gap



PEAKS OR TROUGHS

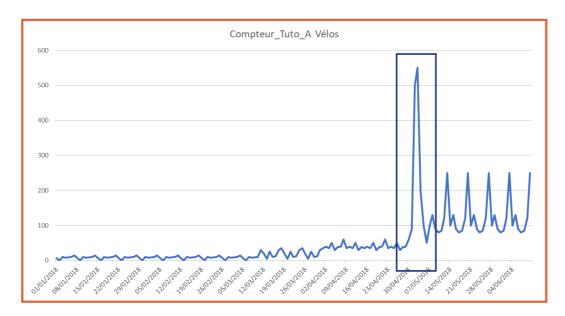
I. Characteristics and visualization

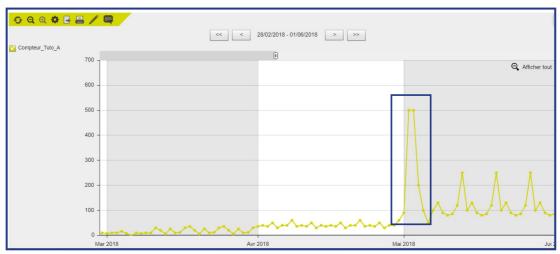
Data peaks are characterised by high (unusual) figures over a short period of time (a few hours to a few days). Figures can be simply high but plausible (500 passages over a day when the norm is 50). But they can also be much higher than attainable figures (e.g. 5,000 passages in one hour, which would be equivalent to 83 passages per second).

On the other hand, troughs are characterised by low (unusually low) figures and over a short period of time (a few hours to a few days).

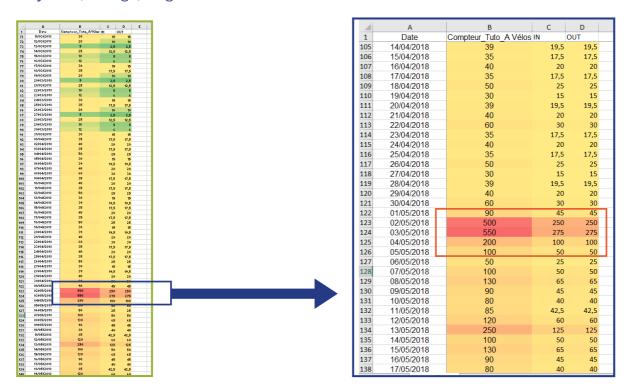
Visualisation of peaks:

In graphical form, the peaks will take the form of a spike-shaped curve that denotes with the usual trend.



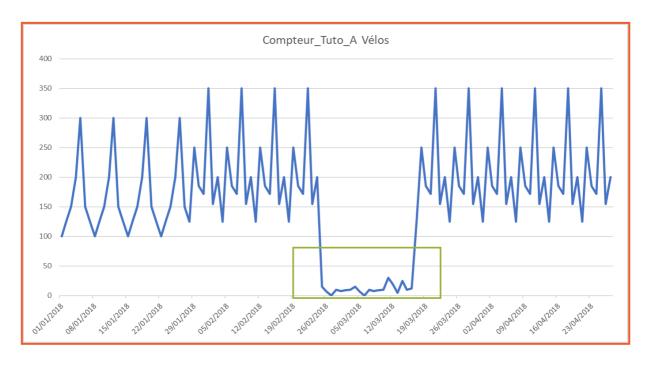


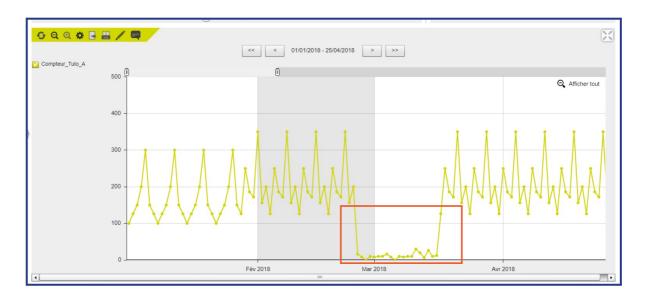
In tabular form (in Eco-Visio or Excel), the values will be higher for the period concerned than the data observed before or after. In Excel, the peaks can be easily visualised using a conditional colour shade formatting. They will appear in red while the other data will be yellow, orange, or green.



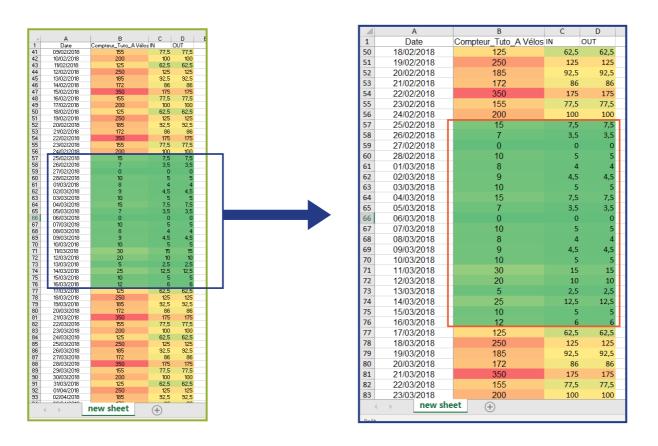
Visualisation of hollows:

In the form of a graph, the hollows will take the form of a pot-shaped curve or a reverse peak and will be different from the usual trend.





In tabular form (in Eco-Visio or Excel), the values will be lower for the period concerned than the data observed before or after. In Excel, the troughs can be easily visualised using a conditional formatting by colour shade. They will appear in green while the other data will be yellow, orange or red.



II. Potential technical or field explanations

Technical:

For non-plausible data: proximity of a disturbing element (high-voltage line, parking over a long period of a vehicle onto the loops, etc.).

Field:

- A peak with plausible data can occur when an event is organised (cycling tour, sports competition, cycling day, etc.).
- A public holiday can also have higher than average attendance.
- A trough with plausible data may occur in the event of bad weather, work on the route with deviations, etc

In all cases, in addition to the material failure, it is important to check with the technical services the work schedule, with the tourist office or the town hall of the area concerned if there was a particular event organised during the period. Sometimes it is also possible to find information on the Internet, as sports or cycling clubs regularly put the calendar of events they organise online, along with the route of the circuits they use. Finally, it may also be useful to check the weather forecast in the event of a very temporary trough. The weather module offered on Eco-Visio allows you to quickly identify whether the weather can be the cause of a specific drop in traffic. In France, it is also easy to find weather reports over a long period of time, particularly on the <u>Bfmtv website</u>.

III. Data processing

If you observe peaks or troughs, in case of technical problems on the counter, you can follow these steps:

- 1. Contact Eco-Counter technical support to understand and solve the problem.
- 2. If a technical failure is proven, a data reconstruction can be considered:
- For a few hours: you can use the automatic reconstruction feature available in Eco-Visio. This calculates for each hour the average of the values of the previous four weeks.
- For several days of data: a reconstruction can be carried out using counters with a high correlation of data (usually nearby counters). This reconstruction can be carried out by you via an import in Excel format, or on request to Eco-Compteur (on quote).

If the observation corresponds to a real event, it is preferable not to modify the data.



INCREASES OR DECREASES

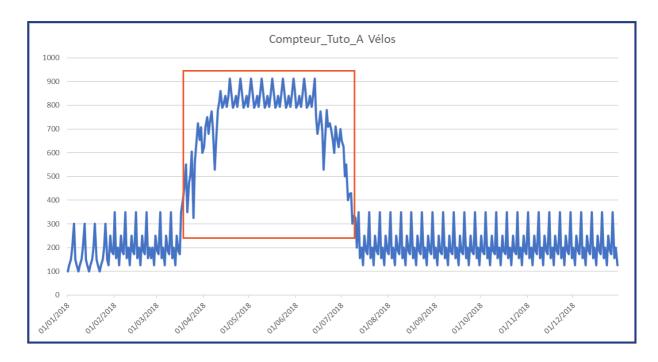
I. Characteristics and visualization

Increases and decreases in counts are characterized by a change in the average trend of the observed daily data. These are rather stable and will be modified within a few days and over a rather long period of time.

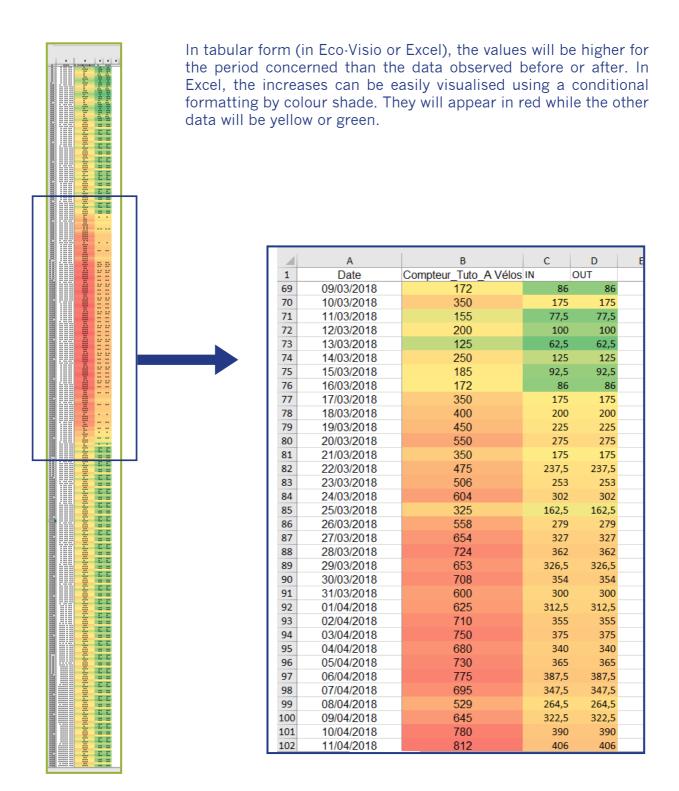
When the data are higher than usual, we speak of an increase. Conversely, when the data are lower, it is called a decrease.

Visualization of increases:

In graphical form, the increases will take the form of a hump-shaped curve.

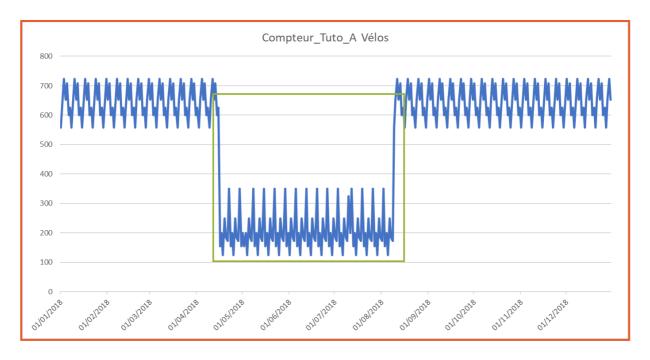






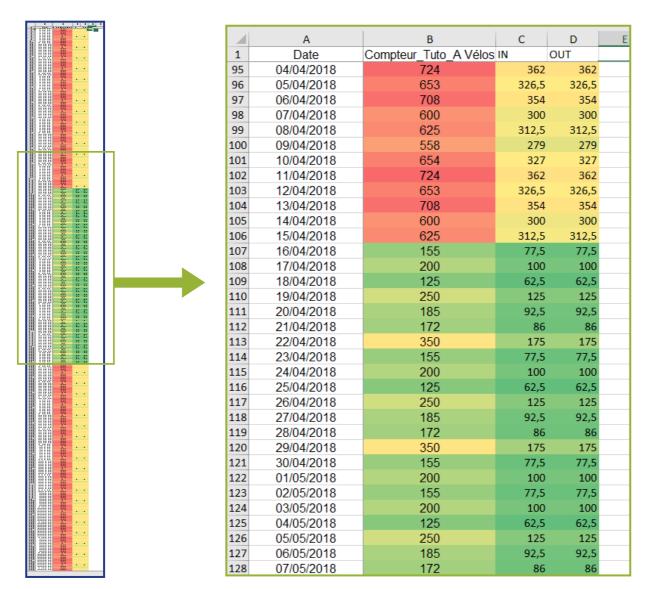
Visualisation of decreases:

In the form of a graph, the decreases will take the form of a trough-shaped curve and will concern a more or less long period of time.





In tabular form (in Eco-Visio or Excel), the values will be lower for the period concerned than the data observed before or after. In Excel, the decreases are easily visualised using a conditional colour shade formatting. They will appear in green while the other data will be yellow, orange, or red.



II. Potential technical or field explanations

Technical:

- In the event of an outlier without recovery of the usual data profile, if no changes or events have occurred in the field, it is advised:
 - A loop or sensor is faulty or has been damaged.
 - In the case of ZELT loops, a surfacing of the coating has been carried out and the coating is too thick. The sensor does not detect the passages.

- When you are not the owner of the counter, in case of an outlier with recovery of the usual data profile, it will be interesting to check that there has not been any technical intervention made on the counter. To check if it is a technical problem, it is advisable to:
 - Check the data by traffic direction when the counter has this option (see next paragraph).
 - Go to the site to check the system and test with a bicycle the correct counting of passages through all the loops.

Field:

• Check whether there has been a temporary change in the surrounding cycle network. This could lead to the postponement or diversion of traffic on the observed site.

III. Data processing

If you notice an increase or a decrease in traffic, in case of technical problems on the counter, you can follow these steps:

- 1. Contact Eco-Counter technical support to understand and solve the problem.
- 2. If a technical failure is proven, a data reconstruction can be considered:
- For a few hours: you can use the automatic reconstruction feature available in Eco-Visio. This calculates for each hour the average of the values of the previous four weeks.
- For several days of data: a reconstruction can be carried out using counters with a high correlation of data (usually nearby counters). This reconstruction can be carried out by you via an import in Excel format, or on request to Eco-Compteur (on quote).

If the observation corresponds to a real event, it is preferable not to modify the data.



INCONSISTENCIES IN THE HOURLY PROFILE

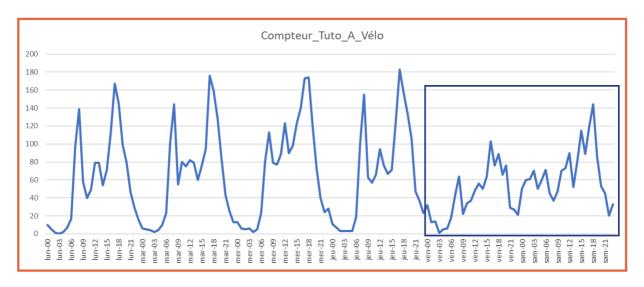
I. Characteristics and visualisations

The hourly data of a counter is naturally distributed in a way that is specific to each site. This distribution constitutes the counter's hourly profile. It is generally stable over the year but can vary significantly in tourist areas, especially during the summer holidays.

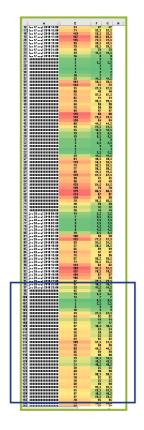
A sudden change in a counter's hourly profile should be closely analysed to check the relevance of the data.

In addition to the analysis of the overall time profile, special attention should be paid to night-time data. Few counting sites record traffic between 10 pm and 6 am (except in big cities). It may be interesting to filter the data to specifically check the data recorded during this period.

In the form of a graph, the usual pattern of the curve on an hourly basis will be changed.







In tabular form (in Excel), when the data are analysed at hourly level and using a conditional formatting by colour shade, the usual distribution shows an alternation of red, orange, and green periods. As soon as the profile is changed, this alternation disappears.

We can find here all the previously mentioned cases: peak, trough, increase, decrease, data at 0, missing data.

It is important to check the hourly data even if they are not used in the published analyses of the data. It is all the more important to check these data as sometimes an anomaly in the hourly counts will not necessarily be visible on the total data but will reflect either a real phenomenon in the field that is important to know, or a technical problem on the counter.

II. Potential technical or field explanations

Technical:

- Missing data: the time change (from summertime to wintertime) frequently results in the creation of an empty time slot at 2:00 a.m.
- Missing data: counters connected to traffic cabinets may show 'missing data' outliers due to maintenance operations on the cabinets.
- In the case of manually read counters, if the reading device is not on time, this can lead to a change of time profile in the data.

Field:

• A one-off event can have an impact on the time profile of a day (cycle race, cycle touring, etc.).

III. Data processing

If you observe inconsistencies in the time profile of your data, you can follow these steps:

- 1. In the case of a manually read counter, check that your reading device is on time.
- 2. Contact Eco-Counter technical support to understand and solve the problem.
- 3. If a technical failure is proven, a data reconstruction can be considered:
- For a few hours: you can use the automatic reconstruction feature available in Eco-Visio. This calculates for each hour the average of the values of the previous

four weeks.

• For several days of data: a reconstruction can be carried out using counters with a high correlation of data (usually nearby counters). This reconstruction can be carried out by you via an import in Excel format, or on request to Eco-Compteur (on quote).

If the observation corresponds to a real event, it is preferable not to modify the data.



INCONSISTENCIES IN THE DISTRIBUTION OF DATA BY COUNTING DIRECTION

I. Characteristics and visualisations

When a counter is equipped to distinguish between flow directions, it is very often the case that the flow distribution between each direction is relatively stable. This may be balanced between the two directions or one direction may have a significantly higher flow rate than the other. But in any case, the distribution is more or less the same whatever the period observed.

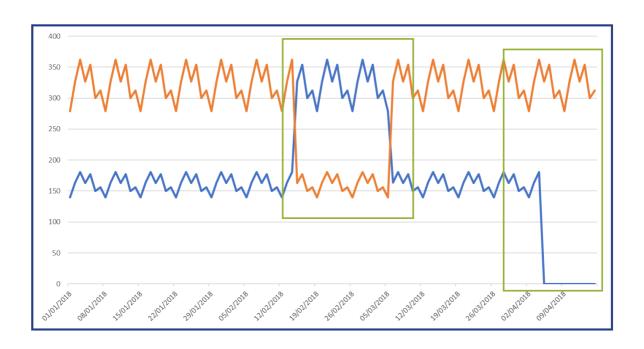
A change in the usual distribution requires special attention. In graphical form, outliers can take different forms.

The graph below shows two different cases:

- A change of predominance in the distribution of flows
- A direction whose frequentation drops sharply to 0.

All the cases mentioned above can be found here: peak, trough, increase, decrease, data at 0, missing data.

It is important to check the data by direction, even if they are not used in the published analyses. It is all the more important to check these data because sometimes an anomaly in the directions of passage will not necessarily be visible in the total data but will reflect either a real phenomenon in the field that is important to know, or a technical problem with the counter.



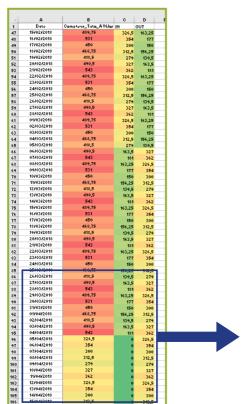


In table form (in Eco-Visio or Excel):

Change of preponderance in the distribution of the flows



Data at 0 on one of the flows



А	В	С	D	
Date	Compteur_Tuto_A Vélos	IN	OUT	
26/03/2018	418,5	139,5	279	
27/03/2018	490,5	163,5	327	
28/03/2018	543	181	362	
29/03/2018	489,75	163,25	326,5	
30/03/2018	531	177	354	
31/03/2018	450	150	300	
01/04/2018	468,75	156,25	312,5	
02/04/2018	418,5	139,5	279	
03/04/2018	490,5	163,5	327	
04/04/2018	543	181	362	
05/04/2018	326,5	0	326,5	
06/04/2018	354	0	354	
07/04/2018	300	0	300	
08/04/2018	312,5	0	312,5	
09/04/2018	279	0	279	
10/04/2018	327	0	327	
11/04/2018	362	0	362	
12/04/2018	326,5	0	326,5	
13/04/2018	354	0	354	
14/04/2018	300	0	300	
15/04/2018	312,5	0	312,5	

II. Potential technical or field explanations

Technical:

- In the case of a 'multi' counter where the direction detection is done by the pyro sensor, it can happen that one of the pyro cells is obstructed by an insect or a leaf. In this case, the total counted by the ZELT loop remains valid. Only the distribution by direction is distorted.
- In the case of a ZELT counter, a malfunction in one of the loops can result in counts, peaks, troughs, or data being reduced to 0.

In both cases, it is necessary to travel to check the operation of the counter on site.

Field:

 Work on part of the track or modification of the network can sometimes explain temporary or permanent changes.

III. Data processing

If you find outliers in the distribution by direction, you can follow these steps:

1. If the observation does not correspond to a real event, contact the Eco-Counter technical support to understand and solve the problem.

2. If a technical failure is proven, a data reconstruction can be considered:

- For a few hours: you can use the automatic reconstruction feature available in Eco-Visio. This calculates for each hour the average of the values of the previous four weeks.
- For several days of data: a reconstruction can be carried out using counters with a high correlation of data (usually nearby counters). This reconstruction can be carried out by you via an import in Excel format, or on request to Eco- Counter (on quote).

If the observation corresponds to a real event, it is generally preferable not to modify the data. In the case of work where a deviation is set up and a counter is placed on it, the data can be corrected using the data from the fixed counter and the deviation.



www.velo-territoires.org

NATIONAL CYCLABLE

info@velo-territoires.org

phone: +33 (0)9 72 56 8505







