CLASSAX-V

Teach-In

Must-Read For All
CLASSAX Teachers

Sensor Line - Gesellschaft für
optoelektronische Sensoren mbH
Carl-Poellath-Str. 19
D-86259 Schrobenhausen
- Germany -
Tel.: +49 (0) 8252 / 8943-0
Fax.:+49 (0) 8252 / 8943-11
Email: sensorline@sensorline.de
www.sensorline.de

1. General

One of the main advantages of the CLASSAX AVC method is its capability to "learn" vehicle classification schemes. Traditional AVC systems require to be explicitly told what each parameter they are able to pick up means in terms of vehicle class. To determine what exactly to tell them requires investigations on the properties of vehicles with respect to the desired classification scheme and the measuring methods applied. Plus, if the measurements performed either during said investigations or during vehicle classification are affected by different systematic errors this difference in "perception" must first be compensated for before the collected data can be transferred to the classification system.

In contrast, the CLASSAX Teach-In process comprises not only vehicle data collection with reference to vehicle classes – the classification method is also implicitly able to extract the meaning of these data. Furthermore, every measurement involved is performed with the same kind of equipment so there is no difference in whatever systematic errors might occur during data collection and classification. This gives CLASSAX great flexibility, it can easily adapt not only to different classification schemes but also e.g. to different typical vehicle dimensions in different countries.

However, it has shown that Teach-In can be done in a more as well as in a less appropriate way. It requires some knowledge about the way CLASSAX works and about the consequences in practice to set up a classification scheme that takes advantage of CLASSAX's full performance.

This document is to give some simplified insight into the CLASSAX method while deriving some hints on how to perform the Teach-In as efficiently as possible.

2. Operation of CLASSAX

2.1 Basic Principle

Essentially, the Teach-In process is nothing but vehicle data collection. The only particular characteristics are that the data are collected with the same measuring equipment which is also used for classification and that the data are not simply stored as they are acquired: CLASSAX computes dynamically their average and standard deviation values, and that is what is stored. (The original data can not be retrieved from these quantities any more so it might be wise to store them first.)

This happens for every parameter used and for every class defined during Teach-In. Nothing else. Average value and standard deviation for every parameter and every class make up the class database.

When required to classify a vehicle CLASSAX supposes that the values of the quantities which were fed into its class database have a Gaussian distribution. If so, average value and standard deviation of a parameter define a bell-shaped curve depicting a measure for the probability that the parameter has a certain value; this probability being the highest at the average value and decreasing symmetrically with growing deviation from it.

Fig. **1** shows how CLASSAX looks at the data from its class database. Every known class yields a quantity on how likely it is for a vehicle of that class that a certain parameter – here the wheel base – has a certain value. The diagram also shows that thus for most values there is one class which yields the maximum probability. This defines a range of wheel base values which can be used to classify the vehicle as being a member of that class.

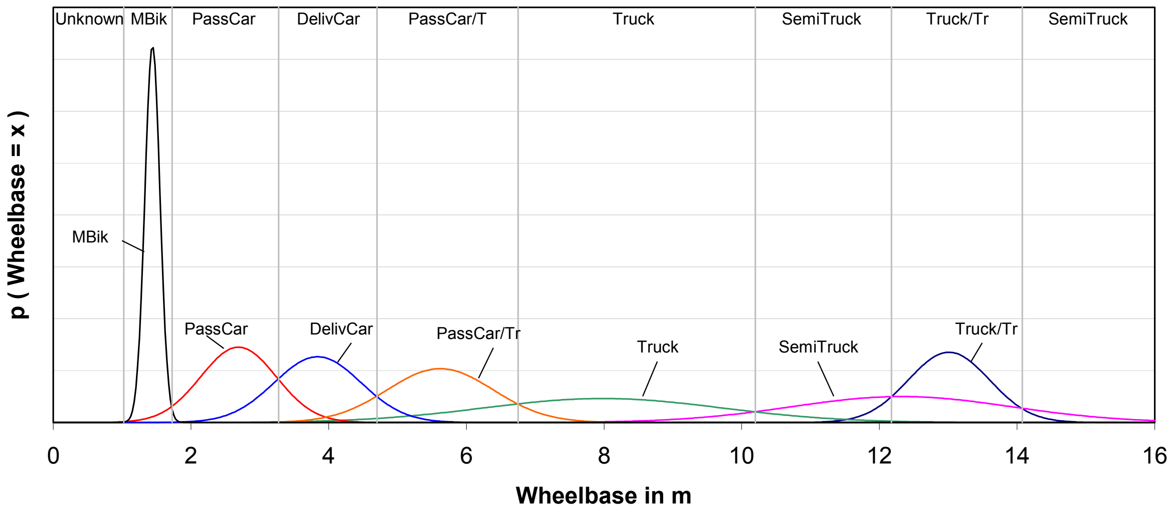

Fig. 1: Interpretation of Class Data for a CLASSAX Parameter

Now this is quite tricky. Actually, to classify a vehicle of given appearance one would ask: "How likely is it that this vehicle belongs to this or that class?" This is difficult to compute. CLASSAX asks: "How likely is it that a member of this or that class looks like this vehicle?" This is almost the same question but the probability is much easier to compute. CLASSAX chooses the class with the maximum probability derived this way.

2.2 Integration of Multiple Parameters

As can also be seen from Fig. **1** every parameter yields an almost complete classification scheme although this does not appear very sophisticated. One may ask how taking other parameters into account can make this scheme smarter instead of just adding contradictory schemes which are all not much smarter.

This can be done by using operations from the theory of probability. If we know a set of parameters plus the probability that each parameter has some particular value we can compute the probability that one parameter has this value AND another parameter has that value by simple multiplication of both probabilities. That means, we do in no way accept the much too simple classification scheme suggested by Fig. **1** but rather continue using the curves of probability it was derived from.

So we now look at a second parameter, let it be the number of axles – better: axle groups. A vehicle of a certain kind needs to be supported at certain characteristic points. This can be done by single axles but also by groups of two or three, so the number of support points is more characteristic than the sheer number of axles used for this purpose.

In the same manner as we looked at the wheel base we can also look at the number of axle groups; the fact that one is a continuous quantity whereas the other can only assume discrete values is in practice of minor importance. We obtain a diagram as shown in Fig. **2**. (The standard deviation is in theory of course zero, but let us assume, there were some few counting errors during data collection – this does not matter, it works as well.)

As can easily be seen, this parameter alone would lead to a classification scheme still more inappropriate than one derived from the wheel base. Only a value of four would point to a single class; all other values would be ambiguous.

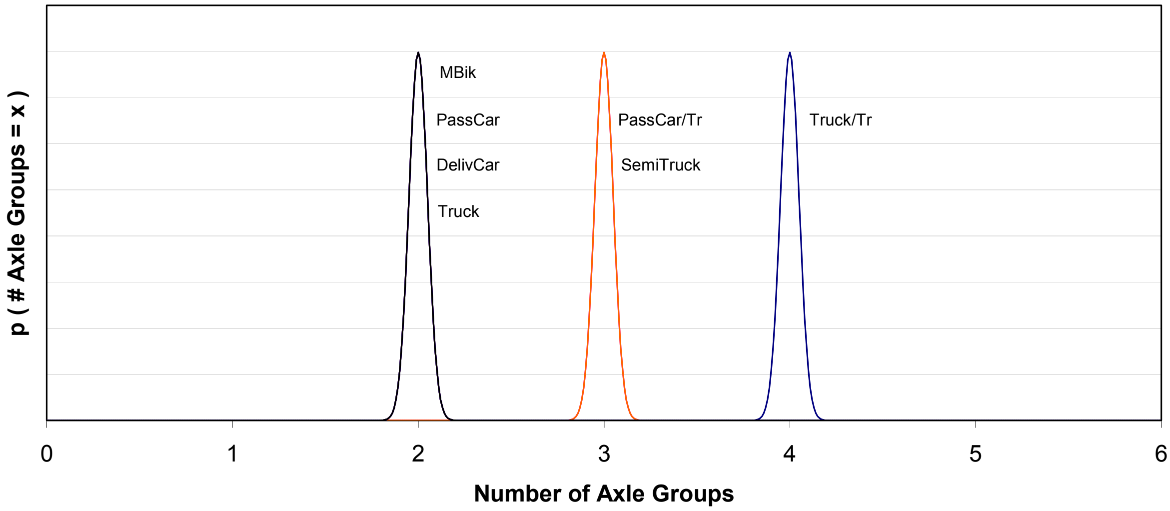

Fig. 2: Interpretation of Class Data for Number of Axle Groups as Second Parameter

However, we do not intend to look at these curves too long, we rather multiply the probabilities as described above. First we multiply each curve from Fig. **1** with the probability that a respective vehicle has two axle groups. We obtain Fig. **3**.

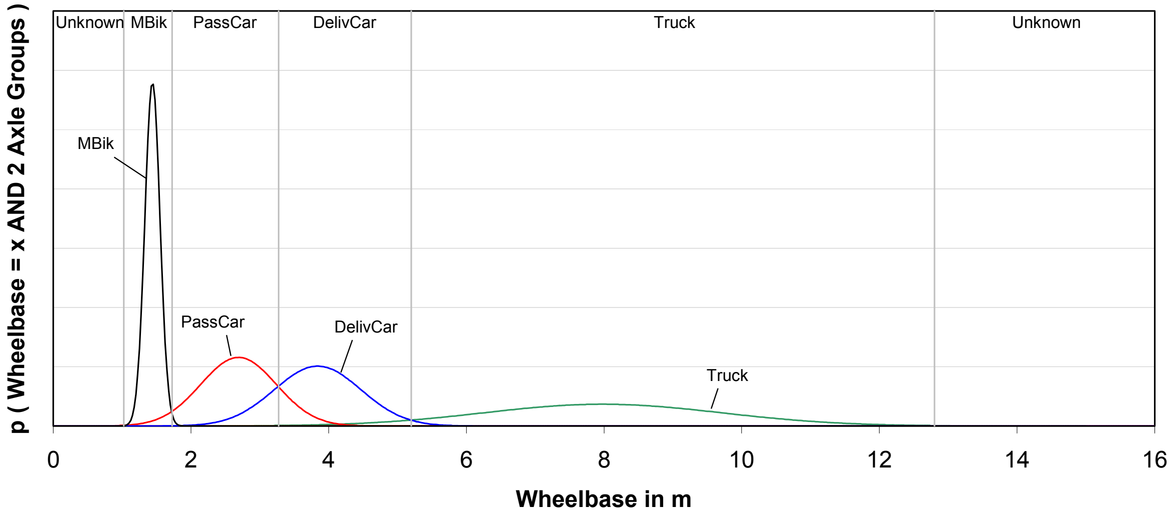

Fig. 3: Class Data for Wheel Base With 2 Axle Groups

The curves for passenger cars with trailers, trucks with trailers and semi trucks have vanished because, as Fig. **2** tells us, the probability that a vehicle of these classes has two axle groups is zero. The same procedure for classes with three and four axle groups supplies us with Fig. **4** and Fig. **5**.

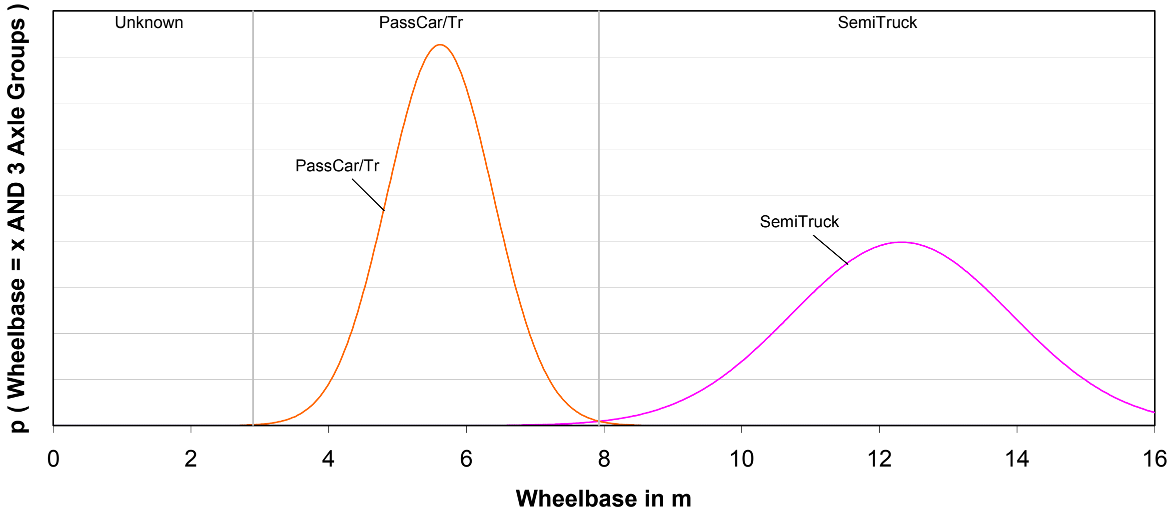

Fig. 4: Class Data for Wheel Base With 3 Axle Groups

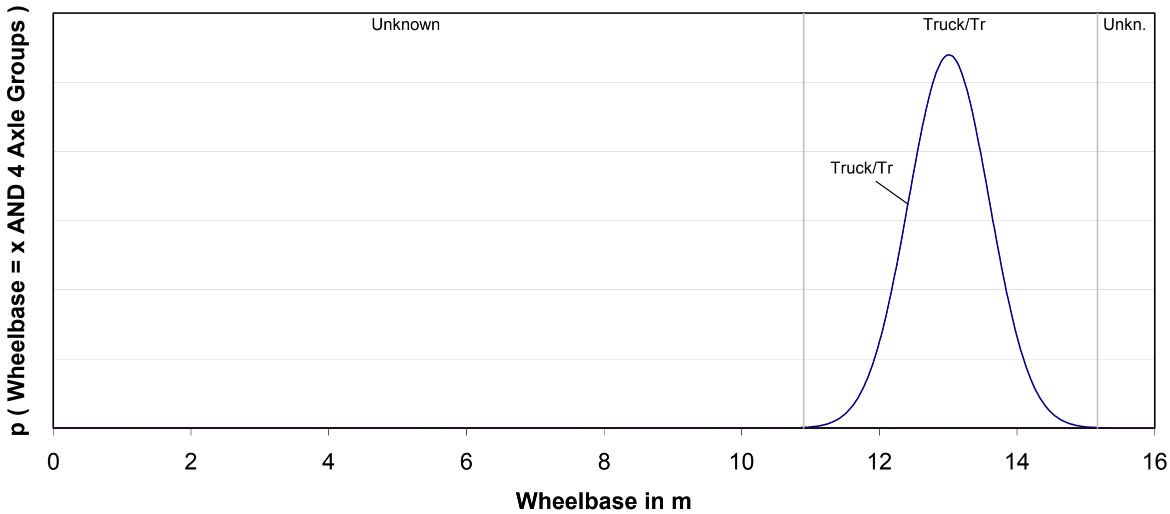

Fig. 5: Class Data for Wheel Base With 4 Axle Groups

These three diagrams make up our new two-parameter classification scheme. It is a whole lot more differentiated than the one contained in Fig **1**: the number of diagrams to be consulted has grown to three according to the possible numbers of axle groups. Moreover, it is also much better: No longer do we have to be afraid of e.g. a long passenger car with trailer to be classified as a truck or a short truck to be taken for a passenger car with trailer. And the funny behaviour of classifying a semi truck below 10m as a truck, up to 12m correctly, then up to 14m as a truck with trailer, and above 14m correctly again has also been eliminated.

Still this classification scheme is far from perfection. We still have overlapping distributions of motorbikes, passenger cars, delivery cars and trucks which all have two axles. This is what additional parameters are meant for: Trucks can be expected to be wider than delivery cars, motorbikes to be much narrower than passenger cars, the latter are still difficult to tell from delivery cars, however.

3. Pitfalls To Avoid

With the knowledge about the operation principle of CLASSAX we can understand some possibilities so set up less appropriate classification schemes by inadequate class assignment during Teach-In.

Let us assume that we have a very simple AVC job: We are only to tell "small" vehicles from "large" ones. Since the vehicles as well as our measuring system remains the same we can easily simulate what might happen by simply replacing the class specifications used to create the examples from the preceding paragraph with "SmallVhcl" for all classes from motorbike to passenger car with trailer and with "LargeVhcl" for everything else.

The teach-in process leads to the wheel base distributions given in Fig **6**.

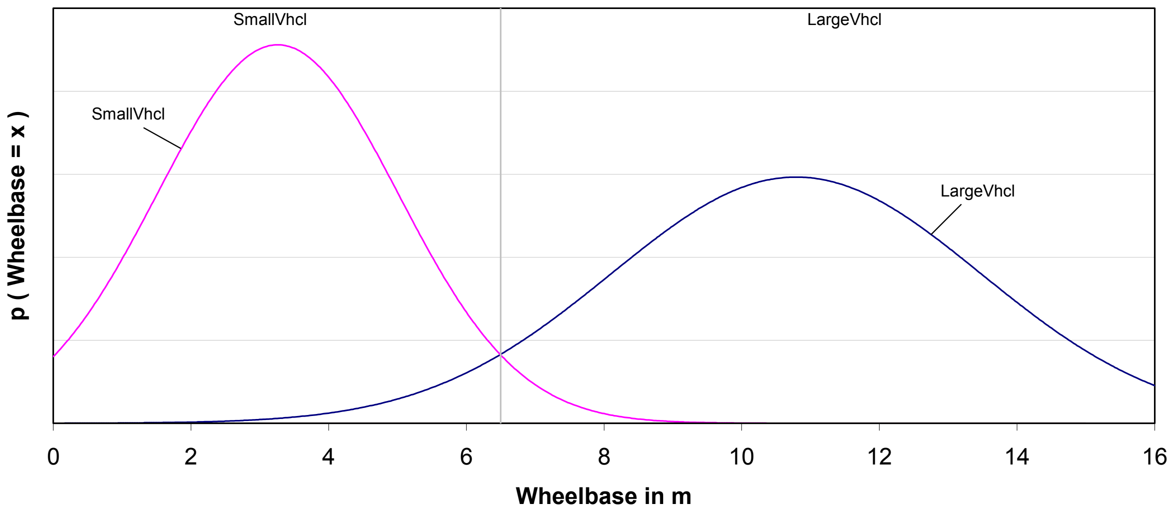

Fig. 6: Wheel Base Distributions with Simplified Classification Scheme

Now that we have only two classes a classification scheme based on the wheel base only would simply define a threshold value at 6.5 meters. It can not at once be seen why it should be unwise to use an algorithm simple like that, and certainly it would classify many vehicles correctly.

But let us have a look at the distributions for the number of axle groups depicted in Fig **7**. From the mathematical point of view this is nonsense because there can not be a probability that a vehicle has something like 2.5 axle groups. In fact distributions of discrete values are not described by Gaussian but by binomial distributions.

However, describing what happened here correctly would not change the fact that it should not have happened: There were vehicles with differing numbers of axle groups entered into common classes. Such must be allowed; we can not provide a particular class for any possible combination of parameter values. On the other hand, in this special case there would have been enough memory space to split both classes into subclasses with different numbers of axle groups.

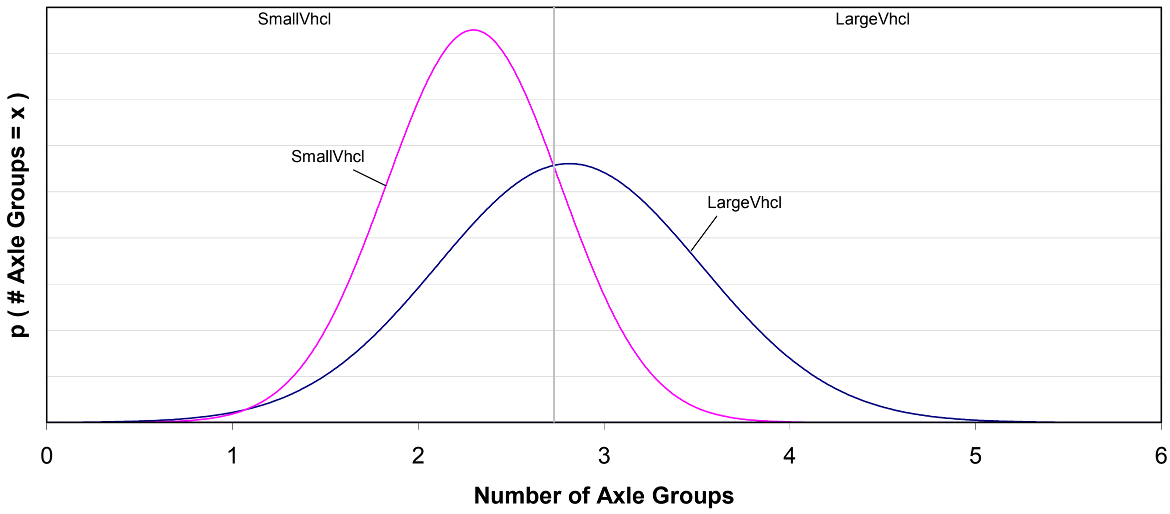

Fig. 7: Number of Axle Groups Distributions with Simplified Classification Scheme

Let us see what happens when we continue the procedure as demonstrated in section 2.2. Again we multiply the curves from Fig. **6** with what the corresponding curve from Fig. **7** tells for two, three and four axle groups. We obtain the results shown in Fig. **8**, **9**, and **10**.

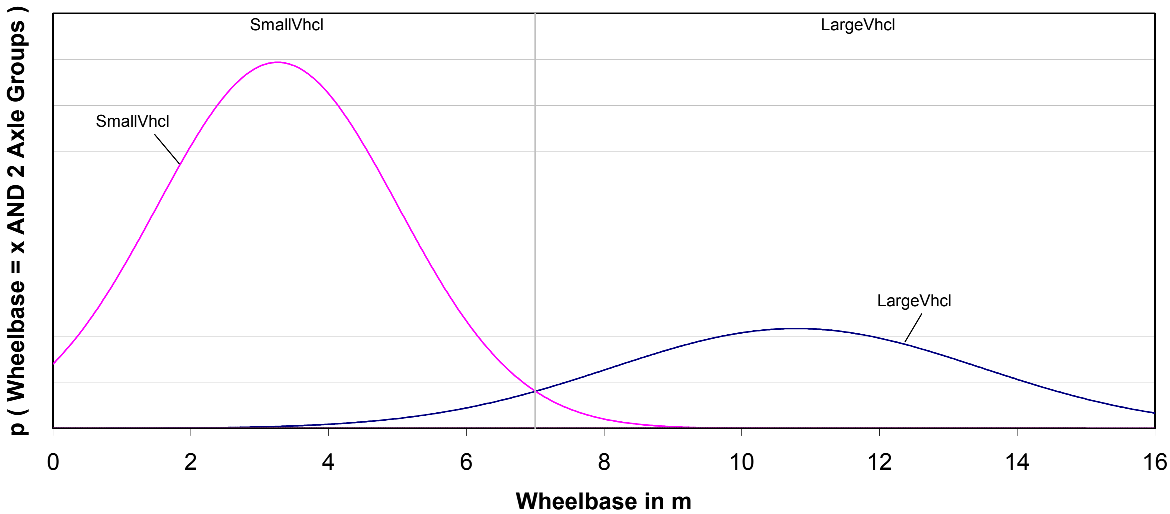

Fig. 8: Wheel Base Distributions for two Axle Groups

Looking at these diagrams we can see that our classification scheme consists of three different wheel base threshold values valid for two, three and four axle groups. One must be aware that with only two classes defined and only the wheel base and the number of axle groups as the only available information this is the most sophisticated classification scheme which can be expected. One can not do more than discriminate between two cases of wheel base at each possible case of axle groups.

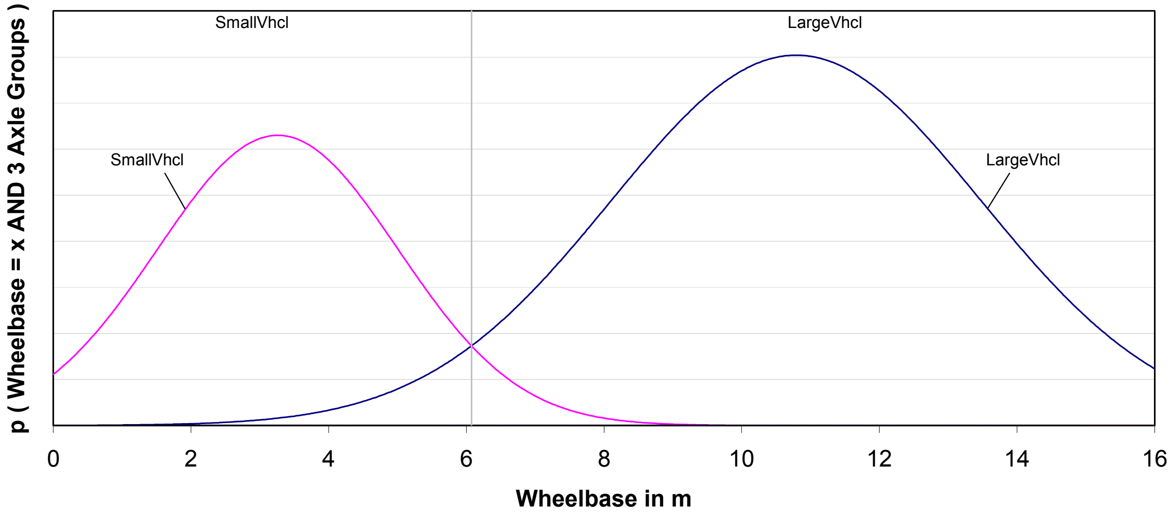

Fig. 9: Wheel Base Distributions for three Axle Groups

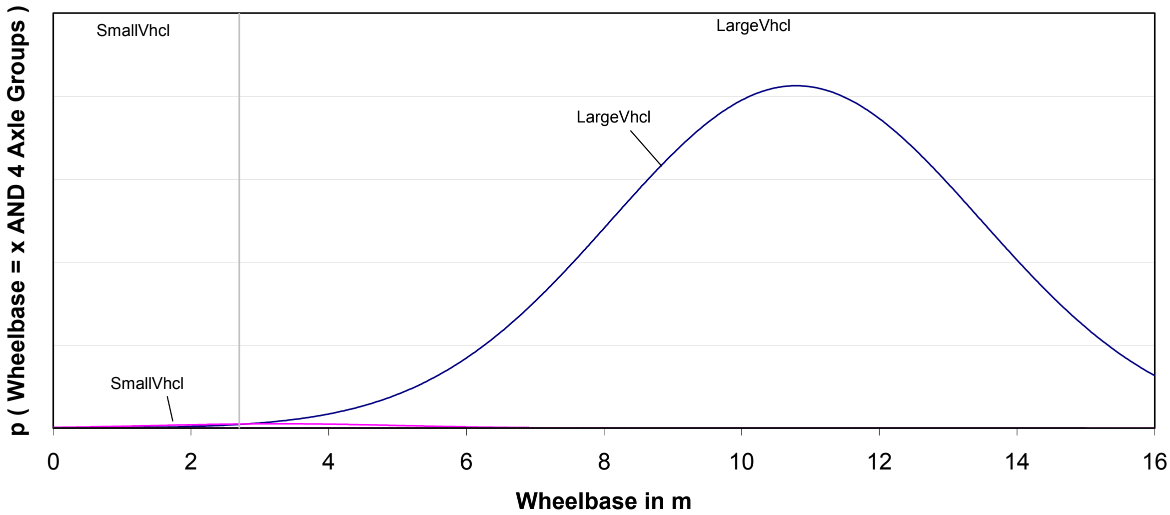

Fig. 10: Wheel Base Distributions for four Axle Groups

As for the quality of this classification scheme we are lucky enough to know from Fig. **1** what vehicles "really" went across our AVC system. Regarding four-axle vehicles we would classify everything above 2.7 meters of wheel base as a large vehicle. This would be all right since there are only large vehicles with four axle groups (trucks with trailers), and their wheel bases start at 11 meters.

Comparing the results for two axle groups with those obtained for the classification scheme from section 2.2 (see Fig. **11** below) we find that CLASSAX now established the wheel base threshold at 7 meters whereas the "real" threshold between delivery cars (small) and trucks (large) would have been at 5.2 meters (dotted line). So although all motorbikes, passenger cars and delivery cars would be classified correctly as small vehicles many shorter trucks would also be classified as small vehicles, and this would not be correct at all.

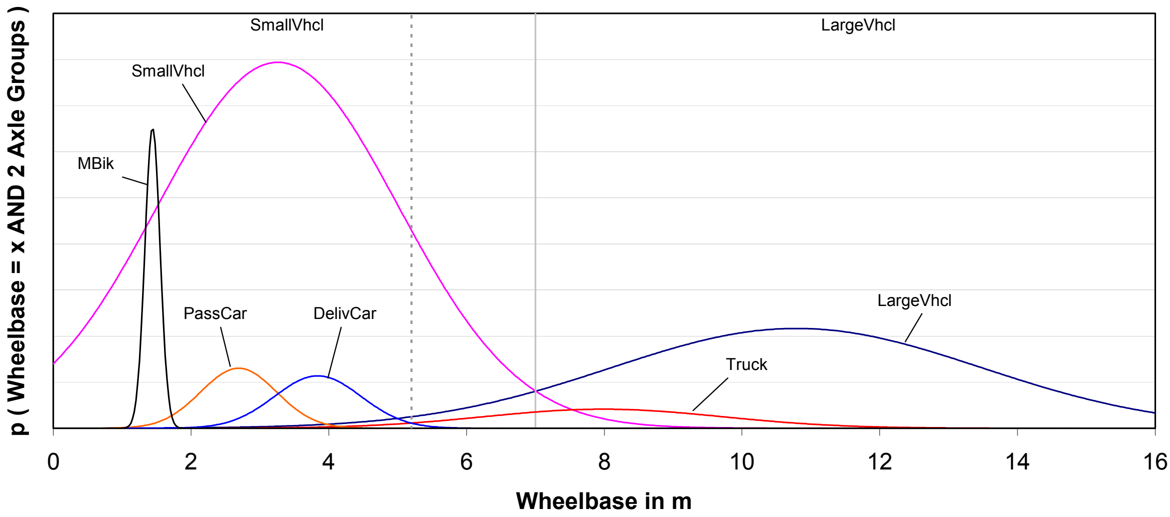

Fig. 11: Comparison of Wheel Base Distributions for two Axle Groups

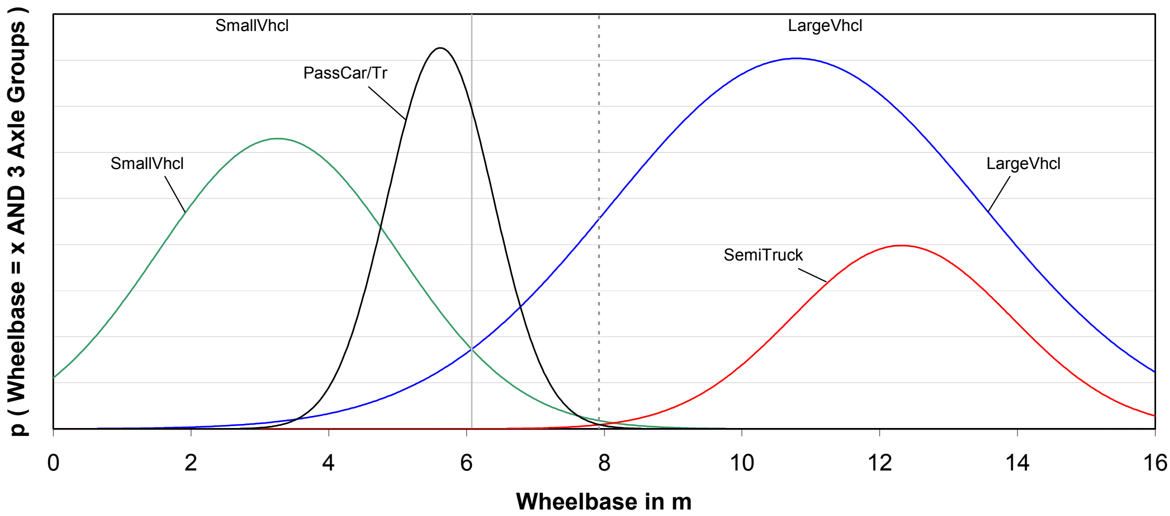

Fig. 12: Comparison of Wheel Base Distributions for three Axle Groups

For vehicles with three axle groups there is also a flaw of this kind. CLASSAX put the wheel base threshold here at 6 meters. However, there are many passenger cars with trailers exceeding this wheel base which would be classified as large vehicles. The correct value would have to be a 8 meters where the wheel bases of the passenger cars with trailers end and those of the semi trucks begin.

These two quite apparent cases of wrong classification were introduced only by the fact that during Teach-In vehicles with different numbers of axle groups were mixed together into common classes. If these vehicles would have been separated into an own class for each number of axle groups the result would have been almost the same as in section 2.2: Large vehicles with 4 axle groups are the same as trucks with trailers, such with 3 axle groups are identical to semi trucks, such with 2 axle groups the same as trucks. Small vehicles with three axle groups are passenger cars with trailers; only the classes motorbike, passenger car and delivery car would have been joined into some common class "small cars with 2 axle groups". The wrong thresholds in the above example would have been set correctly.

4. Conclusion

We learn from the above example that CLASSAX – as everybody else – needs clear specifications to do its work decently. It is not sufficient to tell it for example "Large vehicles may be long but also of medium length and can have any number of axle groups whereas small vehicles are mostly short and have up to three axle groups" as we tried to do. In contrast, telling it something like "Large vehicles are either long with three or four axle groups or of medium length with two axle groups; small vehicles are either short or of medium length but will then have three axle groups" will work.

It is sometimes difficult to comprehend how CLASSAX identifies the correct vehicle class, because this is not done in a "logical" (i.e. "IF...THEN...ELSE") manner. But this does not mean that it can get along without sufficient information. In order to provide this information properly its teacher is in no way released from thinking but must have understood the consequences of feeding particular vehicle data into the Teach-In algorithm plus know some tricks to get CLASSAX working the way he wants.

At least we can derive some general rules:

- A good vehicle class is learned from vehicles which appear similar, i.e. their parameter values remain within small ranges so the standard deviations in the class database remain small.
- If a vehicle class is defined in a way which allows parameters to vary within large ranges one should try to divide this class into subclasses which describe similar looking vehicles, e.g. "commercial vehicles" into "trucks", "semi trucks" etc. This is especially important when parameter ranges overlap with those of other vehicle classes.
- If a "maverick" occurs, e.g. a custom limousine with over-length, do not teach its data into the respective class (here: passenger car) but decide whether to start a new vehicle class or to leave the data unlearned. It is not appropriate to waste memory space for kinds of vehicles that occur twice a year. If on the other hand rare vehicles with deviating characteristics are to be recognized as well an extra class is required to keep the large number of "normal" vehicles from overriding the deviating parameters.
- Never try to teach erroneous measurements into a good class, hoping CLASSAX will be error-tolerant enough to recognize the vehicle anyway. This will spoil good class data. There are, however, some subtleties:
    - Strictly systematic measuring errors are not errors in the sense meant here. With the current sensor setup tire widths can be expected to be measured generally too small. This is all right and no need to worry because it happens always and there will never be a discrepancy confusing CLASSAX.
    - If an error mode is typical for a certain vehicle class this error may of course be treated as a characteristic of this class. In this case, however, the erroneous measurements may not be joined with the good measurements in one class; it is necessary to define a particular class for them.