Frequently Asked Questions
Updated 2006 (TRIM version 3.51)

PREPARING DATA & RUNNING TRIM

*** Which directories are used for input files and output files?**
See menu-item "Options" and page 22 in the manual.

*** Are there any requirements to the data?**
Yes, each site has to contain at least one positive count (value > 0). For time-effect models, each year has to contain at least one positive count. For time-effect models with covariates, each covariate category has to contain at least one positive count for each year. The last two requirements are relaxed for linear trend models: positive counts are only needed for the years selected as change points.

*** Is there any lower limit in the number of sites to be meaningful?**
No. But be careful in using data with a few sites only. The standard errors might be underestimated as models easily fit very good.

*** Is it possible to treat monthly and yearly data simultaneously?**
TRIM cannot estimate models with both monthly and yearly parameters. Such models are necessary for e.g. waterfowl counts. (In the Netherlands, we currently use a combination of Uindex, TRIM and TrendSpotter software to calculate indices for waterfowl).

*** Is it allowed to use TRIM for species that were counted completely?**
This may be the case for some rare species for which e.g. all breeding pairs are known for each year. Such data may still be treated with TRIM. The input data file needed for TRIM contains only one site with the total counts per year. TRIM then assumes the total counts per year being Poisson distributed and calculates standard errors for time totals being the square root of the counts (the same assumption is made when one applies a X2 test on total counts per year).

*** If the sites differ in area, is it necessary to convert the counts into densities?**
No. Differences between sites are taken into account in the site effect in the model. But note that sites with a large area contribute more to the indices than small sites if the number of individuals is larger. See the powerpoint presentation available at www.cbs.nl <http://www.cbs.nl> for a basic understanding of indices.

*** If a new site enters the scheme many years after the start of the scheme, does it mean that all earlier years for that site are missing values for TRIM?**

Yes. One needs to fill in missing values in the data file for all those years for that site.

*** Which files do I need to prepare to run TRIM?**

You only have to prepare a data file (see manual page 19 for specification). You have to make a separate data file for each species rather than one file with all species data. If you have many data to process, it may be useful to compose tcf files too.

*** What is a tcf file?**

TRIM saves the description of a data file in a Trim Command File (tcf). By that, one may work with the previously given data description in a next TRIM session. When also the keyword RUN is included, TRIM runs models without intervention of the user, which is useful if one uses TRIM on a routine basis (see page 23 in the manual).

*** Is it possible to assess indices for all species in one run?**

There are several options to run TRIM for many species in batch mode. One of them is described in the manual (page 23: holding down the shift key while selecting tcf files). Another option is to incorporate TRIM in an automation system based on Access, Delphi etc. Ask Arco van Strien (asin at cbs.nl) for more information.

*** What does the error message "unexpected value site…" means?**

The description of the data file does not correspond with the content of the data file. That may be due to e.g. mistyping the number of years or the number of covariates in the file description dialog box.

*** Is it possible to open output files generated in earlier TRIM runs?**

Not in TRIM. One should use a standard editor to read the earlier output files. Also graphs of earlier runs cannot be displayed in TRIM. Graphs of the current run can be copied to clipboard.

*** What is an ocv file?**

TRIM produces ocv output files (Output CoVariances). These files are necessary output in the framework of the Pan-European common bird monitoring project to assess European trends of birds (www.ebcc.info <http://www.ebcc.info>). These files make it possible to combine indices from many countries to produce results as if the raw data were available. These files are of no use to others and may be deleted.

CHOOSING A MODEL IN TRIM

*** What is the standard approach to compute indices?**

Use the time effects model with serial correlation and overdispersion switched on. This estimates

yearly parameters (in case the time points are years) and produces yearly indices.

*** What is the use of the linear trend model?**

Choose the linear model if the time effects model is not possible due to model estimation problems (see next FAQ's about error messages). Such problems are usually due to scarce data in one or more years. If a time effects model is not possible, then try to run a linear model with all change points selected, except the one or few years that have caused the difficulty. The results will then approximate the time effects model. If all years are selected as change points, the linear trend model is equivalent to the time effects model. It is important to understand that a linear trend model also produces indices for each year, but not based on yearly parameters as in the time effects model. Instead, the linear trend uses the trend between change points to compute the indices. A linear model with change points may also be useful to test trends before and after particular change points. Options in TRIM are (1) to test trends before and after a priori selected change points or (2) to let TRIM search for the substantial change points by using a stepwise procedure. But be careful in using a linear trend model without any change point selected (the default value)! This TRIM-model assumes a straight-line relationship between (the log of) counts and years over the whole period studied, which is often unrealistic. As a result, the imputations will often be of poor quality.

*** What is the use of the "no time effects model" ?**

It is of no use, except for model selection: it is meant to compare the fit with other models. This model produces indices that are all equal to the base year.

*** What is the meaning of the stepwise procedure to select change points?**

The stepwise procedure implies that one is fitting a model based on important change points only rather than on all time effects, as in the time effects model.

*** What do error messages mean as "model error: zero observations for change point at time 3"?**

The time effects model requires that positive count data are available for each year (see FAQ on data requirements). TRIM prompts this error if the model cannot be estimated for this reason. The same error may arise for a linear trend model with a change point selected at time point 3. When the data are missing or sparse for the third year, one should try a linear trend model with all change points selected except the third year.

*** What do error messages mean as "model error: zero observations for change point at time 3, category 1"?**

The model cannot be estimated due to lack of data. In this example, the data file contains no observations in the third year of the time series for the first category of the covariate. One has to choose a linear trend model and select all change points except the third year or to join covariate

categories in order to get a value for each year.

*** What do the error messages “invalid floating point operation”, “singular matrix”, “diverging parameters” mean?**

The model cannot be estimated due to sparse data, a too elaborate model or other reasons. A simpler model should be tried, by removing one or more covariates, covariate categories and/or change points (time points) or by switching off serial correlation. Often TRIM indicates the year that causes the model estimation problem. Remove that year from the change-points-selection list, and run the model again. If you were using the time effect model, you have to switch to the linear trend model to be able to do this. Also set the option to automatically delete change points that cannot be estimated (see options in TRIM).

*** What does the error message “too many parameters” mean?**

TRIM produces this error as one tries try to fit a model with too many parameters i.e. more than 100 parameters (80 in earlier versions). This limit is reached in case of for example a time effects model with 20 years and a covariate with 5 categories. One has to try a simpler model, either by reducing the number of covariate categories or by running a linear trend model with less time points.

*** What is the meaning of the Wald test for significance of deviations from linear trend?**

If this test is significant, it means that the time effects model is a better choice than the linear trend model.

*** Why are the default p-values in the stepwise selection 0.2 and 0.15 instead of 0.05?**

These p-values are used in the process of selection of change points and are conventional values in stepwise procedures. In the end, the significance of changes are tested against the 0.05 level. See section 2.7.2 in the manual for details.

*** By adding a new year to the data, the indices in earlier years change. Is that correct?**

Yes, inclusion of the data of a new year may affect the model estimations for all years of the time series. See the powerpoint presentation available at www.cbs.nl <http://www.cbs.nl> for a basic understanding of the calculation of indices. Generally, the indices in earlier years will hardly change, although sometimes they do change considerably. But one should also take note of the confidence intervals of the indices: the new index values usually remain within the confidence intervals of the indices of the earlier assessment.

*** What is a covariate?**

A covariate describes a property of the site, for example habitat type with categories woodland and farmland. Incorporating a covariate in the model implies that missing counts are estimated (imputed)

for each covariate category separately. Thus, a missing count in a woodland site is estimated from the changes in the other woodland sites only and a missing count in a farmland site is derived from the changes in other farmland sites only. When different trends per habitat are to be expected, this may lead to an improvement of the imputations and of the fit of the model. Both the linear trend model and the time effect model can be extended by covariates.

*** What are the advantages of incorporating covariates in the model?**

Advantages of using a covariate are (1) a model with higher goodness-of-fit, with a higher quality of the imputations and usually lower standard errors of the indices and (2) TRIM tests whether the indices differ between covariate categories; this is often interesting to know. When you use a covariate in the model, TRIM calculates separate indices for each category, as well as an overall index based on the sum of the counts of all sites per year.

*** Why do the weight factors have no effect on the indices?**

One does not only have to set the weighting option, but it is also required to include a covariate in the TRIM model. Else weight factor are part of the site effects only, without influencing all indices. If one wants to weight sites in order to adjust for e.g. oversampling of a particular habitat type (see manual page 33), one has (1) to construct the proper weighting factors per habitat type and include these factors in the data file, (2) to incorporate habitat type as covariate in the data file, (3) to select the weighting option in the model dialog box and (4) to select the covariate option in the model dialog box.

*** How to find the indices per covariate category in the output files?**

Page 31 in the manual explains how to calculate the indices from the parameter estimates listed in the out file. But it is much easier to use the "s" output file (see manual for details); in this file are these indices directly available. There are however no overall slopes per covariate available.

*** Is it possible to name the categories within a covariate?**

No, one can only name the covariate itself in TRIM, e.g. habitat, but not the categories as woodland and farmland. You need to remember where the category numbers stand for.

DEALING WITH MODEL FIT

*** How can you judge if the model fits?**

The fit of the model is tested by two tests: the Chi-square test and the Likelihood Ratio or Deviance test (see manual page 15). Usually, the results of these tests are alomst similar. If the p-value of (one of) these tests is below 0.05, the model is rejected.

*** What if the model does not fit?**

In case the model is rejected, try to find a better model by incorporating covariates that describe the differences in change between sites adequately. You may also compare models using Akaike's Information Criterion (AIC): a lower AIC value implies a better fit (see manual page 15). But what if a better model cannot be found? Please read the following questions.

*** Is lack-of-fit of the model relevant in case there are few missing counts?**

Hardly, although it might be worth trying to find a better model; the reward may be indices with smaller standard errors.

*** Which proportion of missing values in the data is allowed?**

It is frequently recommended not using data with more than 50% missing counts, and sometimes even not more than 20%. But a rule of thumb on the proportion of missing counts tolerated is hard to give, because this depends on the fit of the statistical model applied. The more missing counts are present in the data, the more one relies on the statistical model to estimate (impute) missing counts. If the model fits, the imputed values are expected to be close to the observed values, so the higher the proportion of missing values is allowed to be. TRIM allows the user to fit different models in order to find a proper model, with better indices and often smaller standard errors as a result. Some TRIM users have suggested deleting sites with many missing values in order to decrease the overall percentage of missing values in the data. That is a misconception; don't do that.

*** Are indices reliable even if the model does not fit?**

Generally, yes. In case of lack-of-fit, the quality of the imputations and the indices may be limited. But generally a limited quality of the indices is expressed in bigger standard errors of the indices. That is because TRIM converts any lack-of-fit into higher standard errors (provided the overdispersion is set in TRIM). Thus, the indices are generally reliable, but for a proper interpretation of the results, also take into account the standard errors of the indices. In addition, the overall slope and the Wald tests remain reliable, even if the model does not fit.

*** What is overdispersion and what is its impact on the indices?**

TRIM assumes the count data to be Poisson distributed. Overdispersion indicates the degree of deviation of Poisson distribution, and influences the standard errors of the indices and other parameters, not the indices itself. A high overdispersion may result from a lack-of-fit of the model which implies that better models might reduce overdispersion. But a high overdispersion could be also be a characteristic of the species studied, such as appearing in flocks. Of course, such an overdispersion cannot be reduced by searching for better models.

*** What if the data are not Poisson distributed?**

That does not have to be a problem. Deviations from Poisson distribution can be taken into account by including overdispersion in the TRIM models. The best strategy is to always incorporate overdispersion in the estimation procedure; else the standard errors might be seriously underestimated. The only exception to switch off overdispersion may be in case of a total census.

*** What is the impact of serial correlation on the indices?**

Serial correlation describes the dependence of counts of successive time-points (years) and can be either positive or negative. Serial correlation has a small effect on the indices, except when there are very few data. Taking into account serial correlation frequently produces larger standard errors. Incorporate it into the model (unless the model cannot be estimated then), otherwise the standard errors are expected to be underestimated.

*** How should one assess the best fitting model in case of overdispersion and/or serial correlation?**

Unfortunately, the Goodness-of-fit tests and AIC are not valid if the counts are not independent. Poisson observations. This hampers the evaluation of the fit of the model in case of substantial overdispersion (say > 3) and serial correlation (say > 0.4). In such cases, one has to rely on the Wald tests to find the best model.

ABOUT LOGLINEAR REGRESSION

*** Does one need to transform the data before the analysis?**

The ordinary statistical analyses, like linear regression, assume data to be normally distributed. To meet that assumption, log transformation may be required. To avoid taking the log of zero, one needs to add a constant, e.g. 1, to all values. In wildlife monitoring data we may have many zero counts per species, especially for rare species that do not occur each year at a particular site. Unfortunately, for data with many zero values, log transformation is not sufficient to meet the assumption of normality and the results may even depend on the magnitude of the constant that is added. The GLM-models (McCullagh & Nelder (Generalized Linear Models; Chapman & Hall 1989) offer a better alternative for the ordinary approach and these models have become a standard approach to analyze count data. In GLM-models, the normality assumption is replaced by the assumption of a distribution of the user's choice, e.g. Poisson or multinomial. To apply these models transformation of raw data is no longer required (see the next question). See also Ter Braak et al. (1994) for some more theoretical considerations.

*** What is the explanation of the term loglinear regression?**

Monitoring data may be viewed as frequencies in contingency tables, here site by year tables. The GLM-models that describe those contingency tables are called loglineair models. A linear model is not

just a model that describes a straight line through the data. It is much broader: a linear model is each linear combination of components. Non-linear components may often be linearized by transformation. The log in the name of loglinear models is because the logarithm enters the contingency table model naturally. If one derives the expected frequencies in the cells of the tables from the margin totals (similar as when applying a X2 test), the model would be:

Expected value in a cell = year effect * site effect

This is converted into a linear model by log transformation:

Log (expected value) = log (year effect) + log (site effect)

This is a simple loglineair model, and different from the ordinary linear regression on log transformed values. A loglinear model is about the log of the EXPECTED value rather than about the count itself. The expected value will rarely be zero; this only happens if a species is found at no sites at all in a particular year. No transformation of data is required; the log is incorporated implicitly in the model. It is important to understand that indices computed by TRIM are based on the sum of the counts of all sites in a year and not on the sum of the logarithm of these counts (as would be the case in ordinary linear regression on log transformed counts). See also the explanation of the calculation of indices in the powerpoint presentation available at www.cbs.nl <http://www.cbs.nl>.

*** What is the explanation of the term Poisson regression?**

The Poisson distribution is the basic distribution for data in the form of counts, taking discrete values 0, 1, 2... Think of applying a X2 test to a contingency table; this test assumes the count data to be Poisson distributed. Therefore, loglineair regression for contingency tables with frequencies is also called Poisson regression. One way of viewing TRIM is to consider it as an advanced X2 test applied to time series of counts.

*** Are missing values imputed before the model is estimated?**

No, the idea is to estimate a model using the observed counts and then to use this model to predict (impute) the missing counts. Indices are then calculated on the basis of a complete dataset with the predicted counts replacing the missing counts.

UNDERSTANDING INDICES AND STANDARD ERRORS

*** TRIM generates model indices as well as imputed indices. Which are to be preferred?**

Model-based indices are calculated from the summation of model predictions of all sites i.e. the model-based time totals. Model predictions per site are based on the statistical model. Imputed values

per site are observed values, plus, for missing counts, model predictions. Imputed indices are calculated from the imputed time totals. Often, model-based indices and imputed indices hardly differ, and standard errors of imputed indices are expected to be close to the standard errors of model indices (see also the section in the TRIM manual on the equality of model-based and imputed indices). But model-based and imputed indices sometimes differ. In favour of the use of model-based indices is that the indices might be somewhat more stable than imputed counts, especially if the model fits reasonably. In favour of imputed indices is that imputed indices stick closer to the counts and may show a more realistic course in time, especially for linear trend models. We recommend using imputed indices.

*** TRIM generates additive and multiplicative parameters. Which are to be preferred?**

These parameters are different descriptions of the same estimates: the additive parameter is the natural logarithm of the multiplicative parameter. So, it does not matter which one to use, although the multiplicative parameters are easier to understand. The multiplicative trend reflects the changes in terms of average percentage change per year. If this trend is equal to 1, then there is no trend. If the trend is e.g. 1.08, then there is an increase of 8% per year. This means: in year 2, the index value will be 1.08, in year 3 1.08 times 1.08 etc. If the trend would be e.g. 0.93, then there is a decrease of 7% per year.

*** How to interpret the standard errors of indices and slopes?**

The standard errors of indices are useful to see whether the index in a particular year differs from the situation in the base year. The indices estimated by TRIM are expected to be normally distributed. Calculate the 95% confidence interval of the index of a particular year by adding 1.96 times its standard error to get the upper limit of the confidence interval. Subtracting this value produces the lower limit. If the confidence interval covers 1, the index is not different from the base year. Else there is a significant difference ($p<0.05$). Take for example the index of year 5 = 0.91 with a standard error of 0.02. The 95 confidence interval then ranges from 0.87 and 0.95, and the index is significantly lower than the base year. In this way, all indices can be tested against the base year. The same interpretation is valid for the standard errors of the trend slopes. If one would take 2.58 times the standard error instead of 1.96, one gets the 99% confidence interval, associated with the p-value 0.01.

*** Is it possible to test the significance of change between two years?**

This is only possible if one of these years is chosen as the base year. By definition, the index of the base year is 1 and its standard errors are zero. If the confidence interval of the other year covers 1, the index is not different from the base year.

*** Which year should be chosen as the base time year?**

Usually the first year of the time series is chosen as the base year in order to test the changes against

the base year. In case the species is not present in the first year, another year should be chosen, e.g. the last year. Also when the first year has sparse data, it may be better to use another year as base year. By definition the standard errors of the index of the base year are zero and all errors associated with the base year index are transferred to the standard errors of the indices of the other years. In case of few data in the base year, the standard errors of all indices are expected to be big. (It is even allowed to search for the base year that leads to the smallest standard errors of the indices).

*** Why has no bootstrapping being applied to assess standard errors of indices?**

TRIM uses analytical methods to calculate standard errors. Bootstrapping could be an alternative to generate standard errors or confidence intervals. We do not expect different results using bootstrapping. But bootstrapping consumes more computer time for large datasets and it is difficult to apply in case of elaborate models with covariates.

UNDERSTANDING OVERALL TREND SLOPE

*** What is the use of the overall trend?**

In addition to yearly indices, it is interesting to know the trend over the whole study period. One option is using the linear trend model in TRIM without any change points. But this model is questionable because it has poor quality imputations (see FAQ about the use of the lineair trend model in TRIM). Therefore, we have developed an estimate of the trend slope for time effects models and linear trend model with change points. This trend slope is the slope of the regression line through the logarithm of the indices. The computation of this slope takes into account the variances and covariances of the indices.

*** Why are the standard errors of overall slopes so small?**

The overall slopes of TRIM usually have smaller standard errors as compared to ordinary linear regression applied to the indices. As a result, trends calculated by TRIM may be significant, whereas those computed by ordinary linear regression are not. This is due to a conceptual difference. In TRIM the overall trend error only depends on the inaccuracy of the yearly indices, without taking into account the fit of the regression line through the indices. That is because we regard the slope as a descriptive parameter that summarizes the changes reflected by the yearly indices. In ordinary linear regression the errors are bigger, because the distances of indices from the regression line are also part of the error.

*** Which overall slope should one use?**

TRIM version 3.5 calculates two different trends: (1) the slope of the regression line based upon model indices and (2) the slope of the regression line based upon imputed indices. We recommend using the overall trend estimate for imputed indices (see FAQ about model indices versus imputed indices). Both

regression lines are straight-line relationships between indices and years, with intercept. No estimate of the intercept itself is given, because the intercept differs per site. In earlier TRIM versions also slopes without intercept were given (through the base time point), but these are questionable from an ecological point-of-view.

*** Is it possible to compute overall slopes per covariate category?**

No, because we have not incorporated this feature in TRIM. The only way to compute overall slopes per covariate category is to run TRIM separately for each covariate category.

*** How to interpret the overall trend?**

The overall trend can be interpreted in terms of significant decrease, stable population numbers etc. We have incorporated a new trend classification in TRIM (which differs from the earlier one in Appendix C of the TRIM manual). For more details see Trend Classification under the Help option in TRIM version 3.5.