

# TOD, NVTherm and TRM3 model calculations: a comparison

Piet Bijl, Maarten A. Hogervorst and J. Mathieu Valetton

TNO Human Factors

P.O. Box 23

3769 ZG Soesterberg

The Netherlands

Phone: +31 346 3567 277, fax: +31 346 353 977

e-mail: [bijl@tm.tno.nl](mailto:bijl@tm.tno.nl), [hogervorst@tm.tno.nl](mailto:hogervorst@tm.tno.nl)

## ABSTRACT

At the moment there are three potential models and associated measures to replace the FLIR92 model and the MRTD as the standard for sensor performance characterization and TA range predictions. These are: 1) the NVTherm model that calculates the MRTD-ID, 2) the TRM3 model that calculates the MTDP and 3) the TOD model that calculates the TOD. The three models are grossly different in theoretical approach. We ran the models for the same set of hypothetical cameras. As default sensor, we used a 'typical' under-sampled Focal Plane Array camera. Then, we independently varied the pre- and post-filter MTF's over a wide range while keeping the sampling frequency fixed so that the cameras ranged between well-sampled and highly under-sampled. The differences in outcome are striking. For example, for a range of sensors around the default the TOD model predicts that performance is determined primarily by the *sampling* frequency, while NVTherm predicts that pre- and post-filter *blur* dominate in this region. The results with TRM3 are surprising: the model predicts that *increasing* image information content by applying microscan to the default sensor, *decreases* predicted TA performance. The origin lies in the use of the periodic four-bar test pattern and the definition of the MTDP. In the low contrast region, the MTDP and the TOD are similar but the MRTD-ID deviates from these two. The MTDP and MRTD-ID are different even in the well-sampled region where they both should be equivalent to the conventional MRTD. The study shows that the choice of a model has a large impact on sensor design decisions or trade-offs between sensors. The TOD method is the most general of the three approaches and has the strongest basis. Until now, validation of the model predictions is very limited. Therefore, a joint TA performance study with simulated sensors (e.g. those used in the present study) and target contrast as variables would be very useful.

Keywords: Electro-Optical system performance, TOD, NVTherm, MTF squeeze model, MRTD, TRM3, MTDP, target acquisition

## 1. INTRODUCTION

Since it became apparent that the classical MRTD (Minimum Resolvable Temperature Difference)<sup>1</sup> and the FLIR92 model (which calculates the MRTD) are unable to adequately predict TA (Target Acquisition) performance with the newest generation of (usually under-sampled) thermal imagers, several alternative Sensor Performance Measures and models have been proposed. Currently, the three most promising are: the NVTherm system performance model<sup>2</sup> developed at NVESD (US), the TOD methodology<sup>3</sup> developed at TNO Human Factors (NL), and the MTDP/TRM3 approach<sup>4</sup> developed at FOM (Ge). All three calculate or measure a Sensor Performance Curve for both well-sampled and under-sampled systems from which TA ranges can be predicted.

The NVTherm<sup>2</sup> model has evolved from FLIR92 and uses a linear systems approach. Under-sampling (which means that spatial frequencies higher than half the sampling frequency arrive at the detector array) and imperfect image reconstruction after sampling generate so-called *spurious response* (e.g. aliasing). These effects are highly non-linear and cannot be described directly with a linear systems approach. In NVTherm, it is assumed that spurious response (SR) degrades TA performance. A mathematical expression of the SR capacity as a function of the sensor characteristics was derived and the relationship between SR and performance degradation was assessed experimentally in a number of target identification and recognition experiments with simulated sensors<sup>5,6</sup>. In order to tune the model predictions to the measured performance data, this degradation is incorporated in the model as a *contraction* in the system MTF by a certain factor: the MTF squeeze approach. Two new types

of MRTD were defined from which TA ranges can be predicted: the MRTD-ID for identification and the MRTD-Rec for recognition (in the present version, these are equivalent). For well-sampled sensors, the squeeze factor is 1 and the MRTD-ID and MRTD-Rec are identical to the MRTD. Unfortunately, until now no experimental method exists to actually measure the MRTD-ID and the MRTD-Rec for an under-sampled sensor. The reason is, that in a real sensor sampling cannot be replaced by a contraction of the system MTF; only its effect on task performance for certain classes of targets or test patterns.

The TOD<sup>3</sup> (Triangle Orientation Discrimination) methodology uses a quite different approach. Basically, the TOD is a solid laboratory test measure that is more representative for real target acquisition and is less subjective than the MRTD. This is achieved as follows. First, an alternative non-periodic test pattern (an equilateral triangle in four possible orientations) is used that has a closer similarity to target features than the periodic four-bar pattern. Second, a forced-choice test procedure is used. This procedure is not only very similar to the real target acquisition process, it also yields results that are independent of the subjective decision criterion of the observer. The relationship with real target acquisition was validated in several experiments<sup>7,8</sup>. Recently a TOD sensor model<sup>9</sup> was developed that explicitly takes into account non-linear operations such as sampling. It calculates a TOD curve by end-to-end simulation: a sensor model calculates the images of a set of input test patterns. These images are fed into a visual system model which decides the most probable orientation of the test patterns, just like a human observer would. An interesting feature is that the two stages (sensor and visual system model) can be used independently. For example, an 'objective' TOD for a real sensor can be measured using the vision model instead of a human observer. The model makes use of NVTherm input files and pre- and post-filter MTF's calculated by NVTherm which facilitates comparison of the two models.

The MTDP<sup>4</sup> (Minimum Temperature Difference Perceived) was developed from the classical MRTD laboratory test method with minimal changes. The only differences in the procedure are that (1) the observer is allowed to optimise the phase of the four-bar test pattern with respect to the sampling array of the camera, and (2) he has to be able to resolve the bars (instead of the 4 bars, as stated in the MRTD method). For well-sampled sensors, the MTDP is identical to the MRTD. For under-sampled systems, the number of bars in the image is not necessarily 4, but may also be 3 or 2. With the adjustments in the measurement procedure, it is possible to apply the MTDP method to under-sampled imagers at spatial frequencies well above the Nyquist frequency. The TRM3 model predicts the MTDP. The model is not linear, and calculation is performed partly by simulation.

At the moment, each laboratory uses his own favorite Sensor Performance Measure and sensor model. It is not yet known how well these three measures are related. Further, it is not known how well these measures really predict TA performance, under which circumstances they may be applied or which one is best. In validation studies, the measures usually perform well. However, the validation is often very limited and only applied to one of the three measures. For example, NVTherm and MTDP/TRM3 studies have been focussed primarily on the trade-off between blur and sampling, and their validation has been limited to a single point at high contrast where the effects of system noise and the contrast detection threshold of the human eye can be ignored<sup>4,5</sup>! Nevertheless, the models produce a curve that is a relationship between spatial frequency and contrast, and predict the effect of target contrast on range. So it is implicitly assumed that the shape of the MRTD-ID, MRTD-Rec or MTDP curves matches that for real targets. Bijl and Valetton<sup>7,8</sup> showed that the TOD does not only predict the range differences between well-sampled and under-sampled imagers at high thermal contrasts, but also that the shape of the TOD curve (for a CCD camera) matches the contrast vs. range relationship for real targets very well.

In the present study, we investigate the relationship between NVTherm, TOD and TRM3 model calculations. We run the three models for the same set of hypothetical cameras and vary the sensor parameters systematically over a wide range. Also, the TOD vision model is validated against the performance of a real human observer.

The purpose of the study is twofold. First, the results show how well the model outcomes agree or disagree. Second, hands-on experience with each of the three different approaches will provide insight about their applicability and their strong and weak points.

## 2. METHODS

All three sensor models, NVTherm, the TOD sensor model and TRM3, can be subdivided into three parts: pre-filter, sampling and post-filter. The pre- and postfilter are considered to be linear, and are described with an MTF. Typically, the prefilter MTF is determined by two factors: the optics MTF and the size of the detector (which introduces a blur prior to sampling). The postfilter MTF is determined by, for example, an image reconstruction filter, electronic zoom, and display characteristics. System noise is introduced at the detectors. In NVTherm and TRM3, the eye is included as an extra post-filter MTF and the

threshold is set at a certain signal-to-noise ratio. In the TOD model, a sophisticated multi-layer vision model is used (including noise) and a decision routine chooses the most probable input test pattern.

For all models, we take the same ‘typical’ under-sampled Focal Plane Array camera as the default condition (this is the default sensor supplied with the NVTherm model, see section 2.1). Then, we independently vary the pre- and post-filter MTF’s over a wide range while keeping the sampling frequency fixed so that the cameras range between well-sampled and highly under-sampled. Noise is not varied.

Fig. 1 gives examples of the image of these sensors using the TOD model. The input image is a triangle test pattern with size  $S = 2.5 * \text{detector size}$ . The central image is that for the default camera; from this image it can be verified that this sensor is indeed slightly under-sampled (both blur and sampling effects appear). The orientation of the triangle in the central image can easily be determined which means that the test pattern size is above threshold for this camera.

Variation of  $M$  (vertical direction in Fig. 1) means an extension ( $M > 1$ ) or contraction ( $M < 1$ ) of the pre-filter MTF spatial frequency relative to that for the default sensor:  $\text{MTF}_{\text{pre}}(f) = \text{MTF}_{\text{pre,default}}(f/M)$ . Variation of  $N$  (horizontal direction) means an extension ( $N > 1$ ) or contraction ( $N < 1$ ) of the post-filter MTF spatial frequency:  $\text{MTF}_{\text{post}}(f) = \text{MTF}_{\text{post,default}}(f/N)$ .

A contraction of pre- and post-filter MTF yields a well-sampled sensor (upper left corner): the image is mainly determined by the pre- and post-filter blur, there is no spurious response and the dependency of the image on phase (position of the test pattern relative to the detector array) is low. For a contraction by a factor of 0.5 ( $M = N = 0.5$ ), the orientation of the triangle can still be determined (although it is difficult). A stronger contraction ( $M = N = 0.25$ ; top-left image) makes discrimination of the triangle orientation impossible. In the bottom right corner ( $M = N = 2$  or  $M = N = 4$ ) the sensor is highly under-sampled: the image is mainly determined by the sampling array and depends largely on phase, and the amount of spurious response is large. The conditions in the upper right or lower left corner are not very likely to appear.

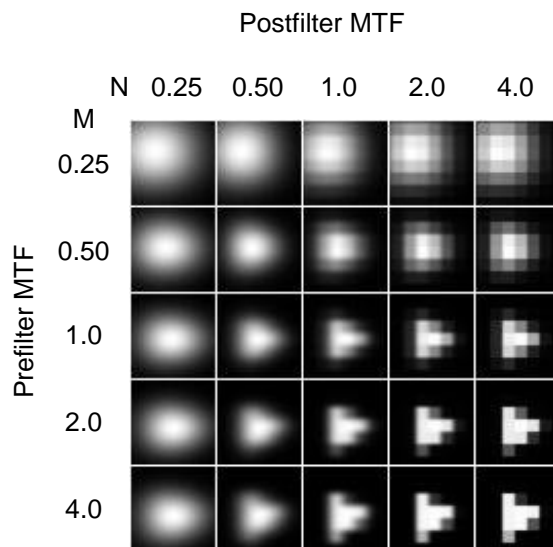


Fig. 1. Image of a triangle test pattern (size  $S = 2.5 * \text{detector size}$ ) with different sensors. The central image is that for a ‘typical’ under-sampled Focal Plane Array camera. Variation of  $M$  (vertical direction) means an extension or contraction of the pre-filter MTF spatial frequency relative to that for the default sensor, variation of  $N$  (horizontal direction) means an extension or contraction of the post-filter MTF spatial frequency. A contraction by a factor of 0.5 ( $M = N = 0.5$ ) yields a well-sampled sensor, for  $M = N = 2$  the sensor is highly under-sampled. See text for details.

## 2.1 NVTHERM CALCULATIONS

NVTherm version dated 27 March 2001<sup>10</sup> was used<sup>1</sup>. As default sensor, the sensor 'default.txt', supplied with the model was taken. This is a simple under-sampled staring array camera with a post-filter that is reasonably tuned to the sampling frequency, no electronic zoom or boost, and a CRT display. Sampling frequency is  $14.66 \text{ mrad}^{-1}$ . NVTherm calculation shows a moderate spurious response: Hor. SR = 0.18, Vert. SR = 0.27; Hor. Out-of-Band SR = 0.06, Vert. Out-of-Band SR = 0.13. The corresponding squeeze factor is 0.93 (which means a contraction of the MRTD-ID spatial frequencies of 0.93 compared to a sensor with the same pre- and post-filter but no sampling). This is in agreement with the relationship derived by Moyer<sup>5</sup>:  $Squeeze = 1 - 0.5 * \text{Out-of-Band SR}$ , and Out-of-Band SR = 0.13 in this case.

A pre-filter MTF extension/contraction by a factor of M was obtained by (1) multiplying the optics aperture by a factor of M, (2) dividing the integration time by  $M^2$  and (3) dividing the detector size by M. Note that the detector pitch (sampling frequency) is not varied. The post-filter MTF was varied by (1) using the calculated CRT MTF as custom MTF and dividing all spatial frequencies by N, and (2) dividing the viewing distance by a factor of N. As output were recorded: horizontal and vertical SR (spurious response), horizontal and vertical out-of-band SR, component and system MTF's, MRTD and MRTD-ID.

Both M and N were varied between 0.25 and 4.0 in steps of a factor of 2 (compare Fig. 1). Calculations could not be obtained for all conditions. For N = 0.25, the program reported: Viewing distance too long (max = 100 cm). For N = 4 and M > 0.5 the program reported: Out-of-Band spurious response too large. For the conditions that were accepted, horizontal and vertical SR varied between 0.00 and 2.36, and horizontal and vertical out-of-band SR varied between 0.00 and 1.02. Increasing M to 8 yielded SR's up to 3.22.

## 2.2 TOD CALCULATIONS

The TOD model<sup>9</sup> uses the same input files as NVTherm, and uses the MTF's calculated by NVTherm. Therefore, the input sensors are identical and NVTherm and TOD results can easily be compared.

Variation of pre- and post-filter MTF is easy since the values are tabulated and all spatial frequencies can be divided by M or N, respectively. Further, the viewing distance has to be divided by N. Calculation of the TOD for the conditions with M and N between 0.25 and 4.0 did not give any calculation problems (in contrast with NVTherm).

For a number of conditions, the sensor part of the model was used to calculate the images on the display (as in Fig. 1), and these images were judged by a real observer instead of the vision model.

## 2.3 TRM3 CALCULATIONS

TRM3 version 2 (July 2000) was used. TRM3 uses input that is different from the NVTherm and TOD input files. Because the sensor we use is simple, conversion to an input file for a comparable sensor in TRM3 is relatively easy. Only minor differences between the sensors remained, like a spectral sensitivity from 3,5-5  $\mu\text{m}$  in TRM3 instead of 3-5  $\mu\text{m}$  in NVTherm and the TOD model.

A pre-filter MTF extension/contraction by a factor of M was obtained in the same way as with NVTherm: by (1) multiplying the optics aperture by a factor of M, (2) dividing the integration time by  $M^2$  and (3) dividing the horizontal and vertical detector size by M. Variation of the post-filter MTF is more complicated. The reason is that some parameters are linked. For example, the signal reconstruction filter (part of the post-filter) is directly linked to the sampling frequency, and the number of tv lines has to be at least the number of detectors. If not, the program produces a fatal error. Extension of the post-filter MTF by a factor N was obtained by (1) using the electronic zoom (N has to be an integer), and (2) multiplying the CRT picture height and number of active lines by N while keeping the eye distance constant (this leaves the MTF's constant but the number of lines is necessary to present all detectors). Contraction of the post-filter was achieved by (1) dividing the viewing distance by a factor of N, (2) dividing the CRT spot size by N, and (3) increasing the number of active tv lines by trial and error until  $\text{MTF}_{\text{post}}(f) =$

---

<sup>1</sup> During the study, the new version 6.0 was released. A comparison showed that the outcome of these two models differed by a few percent only.

$MTF_{\text{post,default}}(f/N)$ . Only a limited number of calculations were made using this procedure. As output were recorded: component and system MTF's and the MTDP. In addition to these calculations, TRM3 was also calculated for the default sensor, but now with the sampling frequency doubled. The reason will be explained in section 3.3.

### 3. RESULTS

In this chapter, we will restrict ourselves to the following conditions:  $M=N=0.5$ ,  $M=N=1$ , and  $M=N=2$  (except in some special cases). These are three conditions with a fixed relation between pre- and post-filter blur (along the diagonal in Fig. 1), and represent a variation between blur on the one hand, and sampling on the other. The conditions represent the important transition range between the region where blur dominates performance ( $M=N=0.5$ ) and the region that is dominated by sampling ( $M=N=2$ ). The reasons for this limitation are: 1) not all models give calculations for all input sensors we defined, 2) this limited set already gives very useful information, and 3) showing results for the other relations between pre- and post-filter would complicate the results and distract the attention from this very important data.

The chapter is organized as follows. Section 3.1 shows the results for NVTherm, section 3.2 for the TOD model, and section 3.3 for TRM3. In section 3.4, a comparison between the models will be presented.

#### 3.1 NVTHERM RESULTS

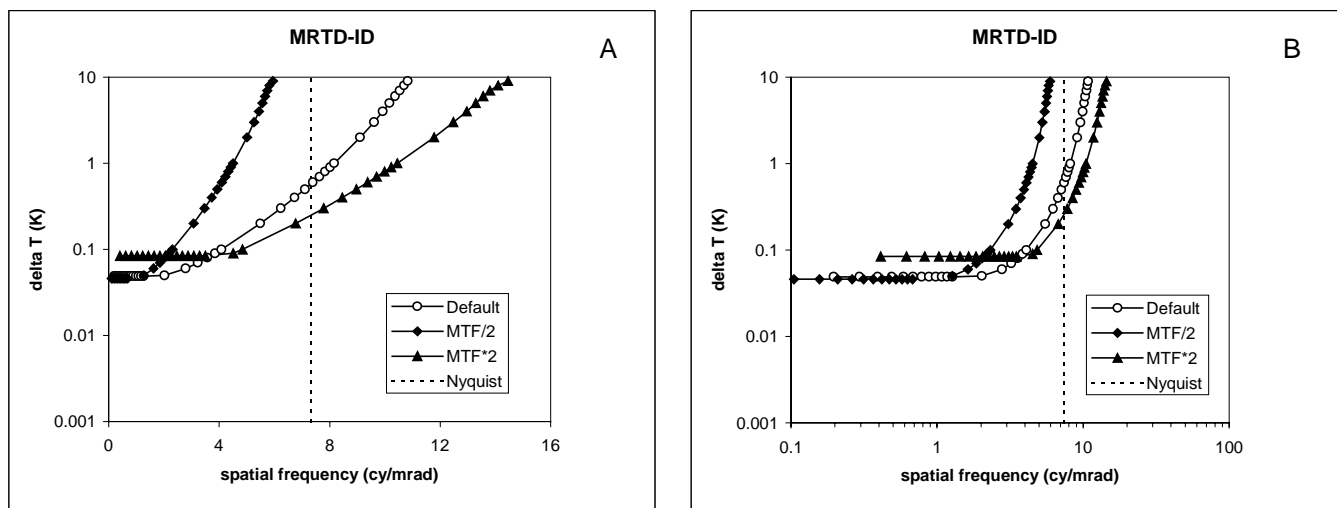


Fig. 2: MRTD-ID calculated with NVTherm. Open circles: default sensor; Diamonds: well-sampled sensor with  $M=N=0.5$ ; Triangles: highly under-sampled imager with  $M=N=2$ . Dashed vertical line: Nyquist frequency. A: conventional lin/log MRTD plot; B: Same data plotted on log/log axes. See text for details.

In Fig. 2, the MRTD-ID's calculated with NVTherm are presented. Open circles represent the default sensor, diamonds the condition with  $M=N=0.5$  (well-sampled sensor, indicated as MTF/2), and triangles the condition with  $M=N=2$  (highly under-sampled imager, indicated as MTF\*2). The dashed vertical line represents the Nyquist (half sampling) frequency. Fig. 2A shows the data plotted in a conventional way on a lin/log MRTD plot., Fig. 2B shows the same data on log/log axes, as often used in vision science. On log/log axes, it is easier to compare the shape of the curves for different conditions. We see that:

- 1) The difference between the MTF/2 and the default condition is a factor of 0.55 in frequency range, which is close to 0.50. This means that pre- and post-filter MTF dominate the MRTD-ID in this region and sampling has little influence.
- 2) The difference between the default and the MTF\*2 condition is 1.33, which is considerably smaller than 2. This means that in this region the effect of sampling is considerable.

- 3) The shape of the curves is the same for all three conditions: in Fig. 2A there is only a shift in horizontal and/or vertical direction. This can be understood since the model only consists of MTF's which differ by a certain factor: either N and M, or the squeeze factor, which is equivalent to a horizontal shift on a logarithmic scale.
- 4) The lower part of the MRTD-ID curves is remarkably constant. This has to do with an interpolation between the horizontal and vertical MRTD-ID. See the Discussion for details.
- 5) For the default and the MTF\*2 condition, the cut-off frequency is far beyond the Nyquist frequency which is the limiting frequency for the conventional MRTD and the FLIR92 model.

### 3.2 TOD RESULTS

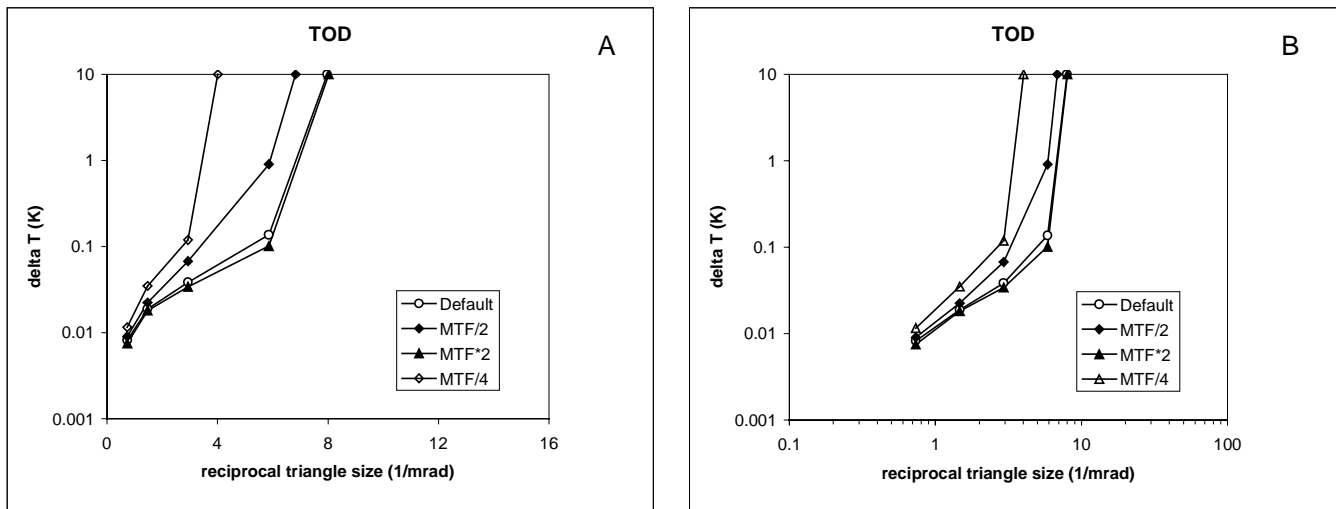


Fig. 3: TOD calculated with the TOD model. Symbols as in Fig. 2. Also plotted: well-sampled sensor with  $M=N=0.25$  (open diamonds). A: lin/log plot with  $1/\text{triangle size}$  (in  $\text{mrad}^{-1}$ ) on the ordinate; B: log/log plot. See text for more details.

In Fig. 3, the TOD model calculations are presented. Symbols as in Fig. 2. Also plotted is the condition with  $M=N=0.25$  (open diamonds) indicated as MTF/4. Fig. 3A shows the data plotted on a lin/log plot, with  $1/\text{triangle size}$  (in  $\text{mrad}^{-1}$ ) on the ordinate, Fig. 3B shows the same data plotted on log/log axes. We see that:

- 1) There is only a small difference between the MTF/2 and the default condition: a factor of 0.85 along the ordinate. The curves for the default and the MTF\*2 condition almost match. This means that sampling and not pre- and post-filter blur is the most important factor in the region between MTF/2 and MTF\*2. This result is largely different from the predictions made by NVTherm!
- 2) The difference between the MTF/4 and the MTF/2 condition is a factor of 0.6. In this region, pre- and post-filter MTF dominate the TOD and sampling has little influence. This is similar to the behaviour expected with NVTherm.
- 3) The shape of all curves is approximately the same: a horizontal shift in Fig. 3B is sufficient to have all data points in a single curve.
- 4) In Fig. 3B the slope of the lower part of the TOD curves equals 1. This means that the threshold is directly proportional to  $1/\text{test pattern size}$ .

In order to validate the vision module in the TOD model, we repeated the acuity measurements (i.e. the cut-off reciprocal triangle size in  $\text{mrad}^{-1}$  at high contrast) for the 4 conditions in Fig. 3 with a human observer viewing the simulated images. In Fig. 4, acuity thresholds for observer PB are plotted against the results with the vision model. The agreement is excellent: the human observer is about 10% better than the model, but this relationship holds for all conditions. We conclude that:

- 5) The vision module in the TOD model is able to predict human performance for largely different types of image degradation (see e.g. Fig. 1): blur, sampling, and combinations of these two.
- 6) The largely different behaviour between NVTherm and the TOD model in the region from MTF/2 and MTF\*2 cannot be attributed to prediction errors of the vision model.

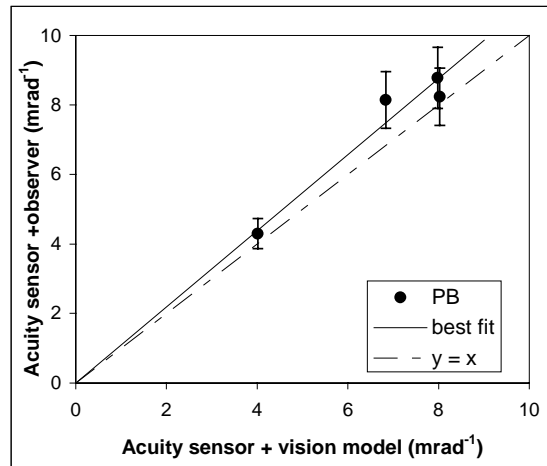


Fig. 4: Acuity thresholds for observer PB viewing the simulated sensor images against the results with the vision model. The agreement is excellent: the human observer is about 10% better than vision module for all conditions. This means that the vision model is able to predict human performance for largely different types of image degradation: blur, sampling, and combinations of these two.

### 3.3 TRM3 RESULTS

In Fig. 5, the MTDP's calculated with TRM3 are presented. Symbols and axes as in Fig. 2 Also plotted is a condition with the default sensor but now with 2 by 2 times rectangular microscan (filled circles, indicated as 2\* $\mu$ scan). We see that:

- 1) There is an unexpectedly large improvement from the MTF/2 to the default condition: when increasing the MTF's by a factor of 2, then MTDP spatial frequency (at least at high spatial frequencies) increases by a factor of 2.4 (i.e.  $> 2$ )!
- 2) The difference between the MTF/2 and the default plus microscan condition (filled circles) is about a factor of 2. This is expected: the spatial properties of all parts of the sensor (both the MTF's and the sampling frequency) are improved by a factor of 2. Only the noise level remains the same, which explains why the curves coincide at low spatial frequencies.
- 3) Findings 2 and 3 together lead to a remarkable conclusion which can be verified in Fig. 5: if the sampling frequency of the default sensor is *increased* (for example by scanning or microscanning), then the MTDP of the sensor is *decreased* (and thus also the predicted TA performance)! This finding is not predicted by the two other models and will be explained in the Discussion.
- 4) The difference between the default and the MTF\*2 condition (highly under-sampled) at low and high contrasts is relatively small. The same was found for the TOD. In the medium contrast region (between 0.1 and 1 K) the MTF\*2 curve is considerably lower.
- 5) The shape of the curves depends on the amount of under-sampling. For a well-sampled imager, the curve is smooth. For a highly under-sampled imager, the curve almost breaks down into three parts: a linear part with slope 1 at lower spatial frequencies (on log/log axes), a steeper part between the Nyquist frequency and the cut-off frequency, and a vertical asymptote.
- 6) The slope of 1 at the lower part of the MTDP curves in Fig. 5B means that the threshold is directly proportional to spatial frequency. The same result was found for the TOD.

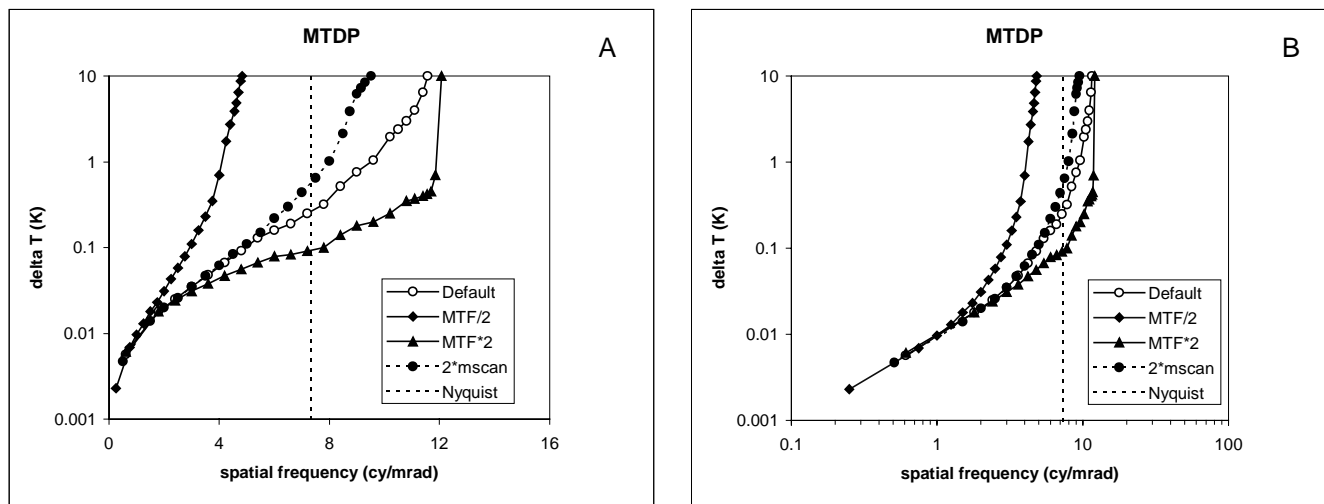


Fig. 5: MTDP calculated with TRM3. Symbols and axes as in Fig. 2. Also plotted: default sensor with 2 by 2 times micro-scan (filled circles). See text for details.

### 3.4 COMPARISON BETWEEN THE MRTD-ID, TOD AND MTDP

In Figs. 6, 7 and 8, the calculated MRTD-ID, TOD and MTDP's (from figs. 2, 3 and 5) are replotted for the well-sampled sensor (MTF/2), the default sensor, and the highly under-sampled sensor (MTF\*2), respectively. Figs. 6A, 7A and 8A show the data plotted on a lin/log plot, with spatial frequency (in cy/mrad) or 1/triangle size (in  $\text{mrad}^{-1}$ ) on the ordinate, Figs. 6B, 7B and 8B show the data on log/log axes. The dashed vertical line represents the Nyquist or half sampling frequency. We see that:

- 1) In the low contrast region there is excellent agreement between the shape for the TOD and MTDP for all sensors. The scaling factor in the horizontal direction is about 0.88 in each condition. This means that the underlying mechanisms of these models yield the same results, *independent* of the sensor type. The absolute value of this factor is of less importance, since it depends on the definition of the triangle size. As mentioned earlier, the behaviour of NVTherm at low spatial frequencies deviates largely from that of the two other models.
- 2) For the well-sampled imager (Fig. 6), the shapes of the three curves are rather different. The MTDP is much steeper than the MRTD-ID and the curves cross, which is surprising because both curves should be equal to the classical MRTD for a well-sampled imager. It means that one of the two models does not correctly predict the MRTD. This could be verified easily by comparing NVTherm and TRM3 predictions with a measured MRTD for an existing scanner.
- 3) For the default under-sampled imager (Fig. 7), the MRTD-ID and the MTDP at medium to high contrasts are more or less the same except for a constant scaling factor in the horizontal direction. When using the MTDP, higher performance ranges will be predicted than with NVTherm. The TOD is steeper than the other two.
- 4) For the highly under-sampled imager (Fig. 8), the shapes of the MRTD-ID and the MTDP at medium contrast are similar, but with a different scaling factor than for the default sensor. At higher contrasts, there is an abrupt cut-off of the MTDP and again the two curves cross. At high contrasts, also the TOD deviates from the other two.

In conclusion, the behaviour of the three models is largely different both as a function of contrast and sensor type, except in some regions. This means that trade-offs between two sensors based on different models may yield a totally different result.



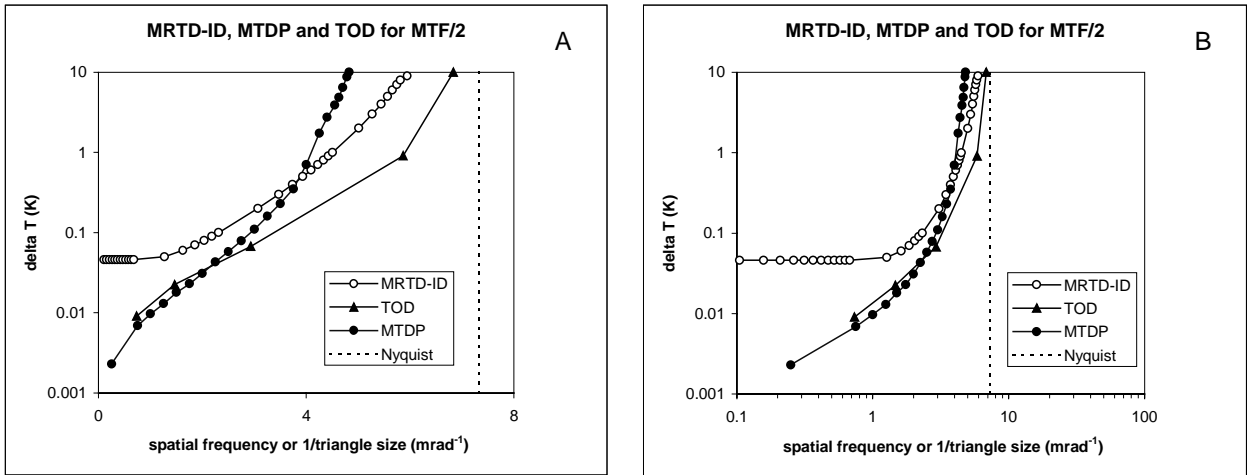


Fig. 6: MRTD-ID, TOD and MTD for the *well-sampled* sensor ( $M = N = 0.5$ , or  $\text{MTF}/2$ ), replotted from Figs. 2, 3 and 5.

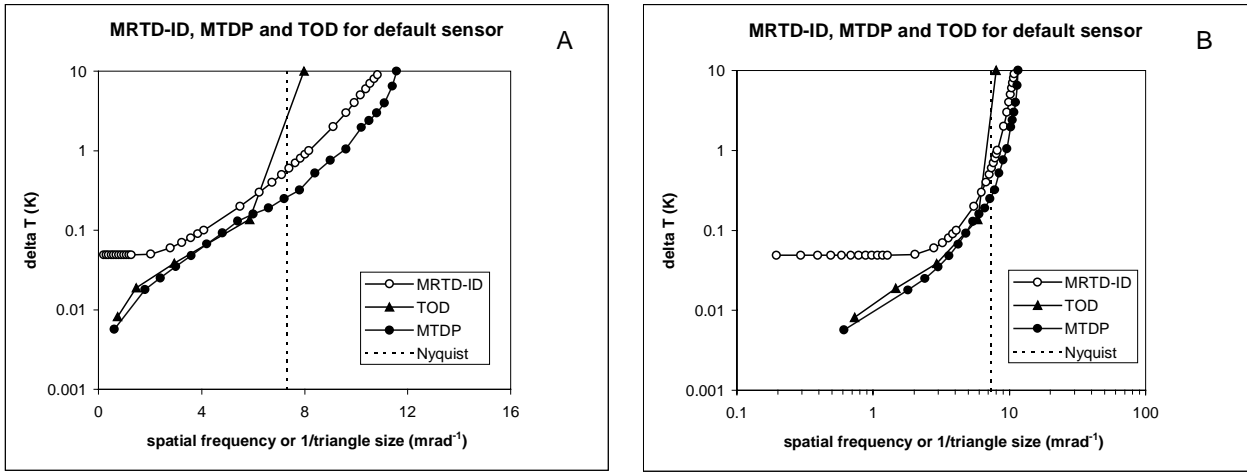


Fig. 7: MRTD-ID, TOD and MTD for the *default* sensor ( $M = N = 1$ ), replotted from Figs. 2, 3 and 5.

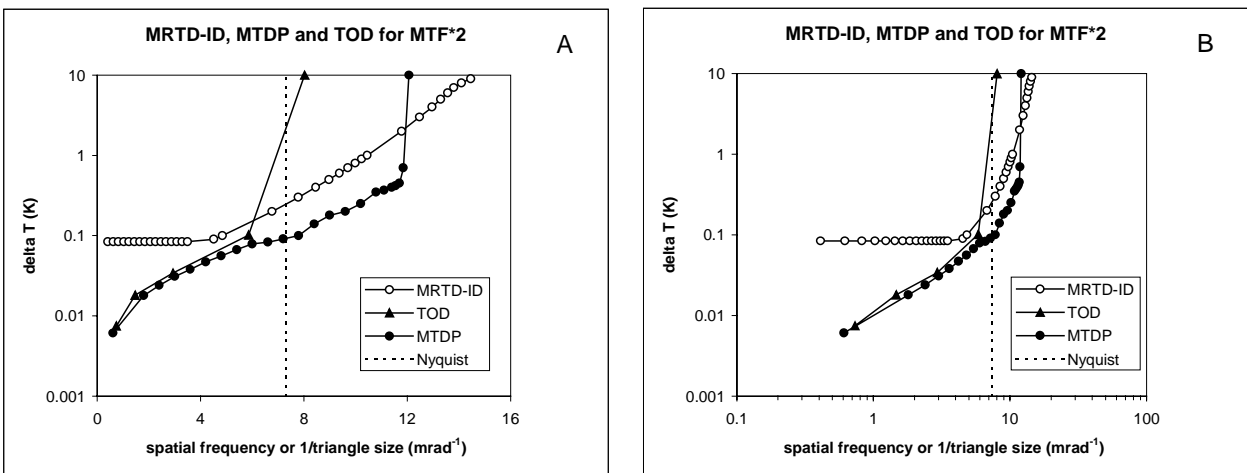


Fig. 8: MRTD-ID, TOD and MTD for the *highly under-sampled* sensor ( $M = N = 2$ , or  $\text{MTF}^*2$ ), replotted from Figs. 2, 3 and 5.

## 4. DISCUSSION

In this section, we will consider the theoretical basis of each of the models in more detail.

The NVTherm model is expected to be able to predict TA performance for a certain range of sensors, because the squeeze factor was tuned to the results of target identification and recognition experiments. The main assumption is a one-to-one relationship between a mathematical measure of non-linear image distortion in a sensor on the one hand, and TA performance degradation on the other. This approach can be useful for discovering systematic effects of pre-filter, sampling and post-filter on TA performance. However, the complex relationship between scene, sensor and human vision is not yet fully understood and a priori it is not known if the assumption is applicable in general. Therefore, numerous TA experiments with human observers are required to obtain a model in which the effects of all sensor parameters are reliably incorporated. As soon as a new sensor property is introduced (e.g. boost or a type of image enhancement), its effect on performance has to be assessed in a new series of TA experiments. Further, the performance of new (unconventional) sensor concepts cannot be predicted by the model. A difference with the classical FLIR92 approach is that the predictions of the current model are based on performance data for real targets which means that it is no longer an MRTD model.

At present, the range of conditions for which NVTherm can reliably be applied, is restricted. As already stated in the Introduction, the relationship between SR and the squeeze factor has only been assessed for the high frequency cut-off of the MRTD curve. Second, the validation studies are limited to a certain range of spurious response (around a normally behaving well- or under-sampled sensor). In Fig. 4 of the study by Moyer<sup>5</sup>, the data from a number of previous experiments are summarized. The maximum SR in these studies is 0.70. In this range, the squeeze factor (obtained by linear regression) decreases from 1 to 0.8 and differences between data points and fit of up to  $\pm 0.10$  occur. The NVTherm program accepts SR values above 3.0 and out-of-band SR's up to 1.0 (see 2.1). This means a considerable extrapolation of a rather inaccurate regression line. In those cases, the user should be informed that predictions are outside the range for which the model has been validated and may be inaccurate. Finally, only simulated sensors with a certain class of pre- and post-filters (a sinc times a Gaussian filter) are used in the TA experiments. In the model, other pre- and post-filters can be selected but these are not validated.

At low spatial frequencies the MRTD-ID is constant (see e.g. Fig. 2). The reason is that the (two-dimensional) MRTD-ID is constructed from the horizontal and vertical MRTD by taking the geometrical average of spatial frequencies at each contrast. These two MRTD's have a (different) local minimum because the eye CSF (Contrast Sensitivity Function) in the model has an optimum spatial frequency. However, taking the geometrical average is only possible above the highest of the two minimum contrasts. Below that spatial frequency, the model takes a constant value for the MRTD-ID. The model can be improved by taking a more appropriate CSF. The CSF depends very much on the way it is measured: when the area (in visual angle) of a sine wave grating is held constant, then indeed a CSF with a minimum threshold contrast is found. On the other hand, if the number of cycles is kept constant (as is the case with the MRTD test pattern), then the contrast threshold increases monotonously with spatial frequency over the entire range<sup>11</sup> (although very slowly below a certain spatial frequency). When this CSF is applied, the geometrical average of the two MRTD's can be taken over a much wider spatial frequency range.

The TRM3 model is based on the perception of a four-bar pattern. Nevertheless, it is very different from the NVTherm approach. The model and the MTDP lab test procedure seem reasonably sound.

Using TRM3 we did a remarkable finding (see section 3.2): if the sampling frequency of the default sensor is *increased* (for example by micro-scanning), this results in a *poorer* MTDP of the sensor! Such a decrease in performance is not found with the two other models, nor is it expected to occur for real targets (because the information content in the image increases). We have the following explanation. Consider an under-sampled sensor. With the MTDP procedure, the observer is allowed to measure at bar spatial frequencies above the Nyquist frequency. After sampling, the number of bars is reduced to 3 or 2 and the spatial frequency is much lower than the input bar frequency. In the second case, consider a sensor with the same pre- and post-filter MTF but a much higher sampling frequency so that the sensor is well-sampled (e.g. by microscanning, scanning or superresolution). After sampling, the number of bars is still 4 and the spatial frequency is not reduced. Thus, with the same input bar frequency, for the well-sampled system the post-filter and the eye get a much higher bar frequency to resolve than for the undersampled sensor and this may result in a much lower contrast and higher threshold to this pattern. As a result, improving your sampling frequency may decrease the MTDP! The effect is strongest if the post-filter is not very good. With a good post-filter and display that can cope with high spatial frequencies the negative effect on the MTDP will be smaller. However, these predictions probably have little to do with real target acquisition and validation is required.

The TOD approach is based on the assumption that recognition of simple spatial test patterns is representative for real TA and the method is close to the real TA process. This seems a relatively reliable approach as long as the interactions between sensor characteristics and human vision in the TA process are not fully understood. The model is the only one that uses a sophisticated human vision module. Basically, the eye models of NVTherm and TRM3 consist of an MTF and a threshold SNR and in principle they are only applicable to the first harmonic of a periodic test pattern. However, under-sampled sensors can distort the image in many different ways. The TOD vision model makes a multi-layer neural representation of the image (including neural noise) and decides which was the most probable of the possible input images. Therefore, it can handle arbitrarily distorted shapes or patterns and its application is not restricted to a certain family of sensor types. In this paper, it was shown that the model very well predicts the effect of different types of degradation (blur, sampling or a combination) on the acuity threshold (Fig. 4). Validation of the vision module at lower contrasts and with different sensors has to be performed.

Since the TOD model is very general it allows performance prediction for new (unconventional) sensor concepts. Application of local or global image enhancement, microscan or an unusual detector shape can be implemented without changing other elements of the model. Further, the vision module is not restricted to certain types of test patterns (equilateral triangles in this case). In principle, the same model can be used to choose the most probable target from a set of real targets (with limitations). Since the sensor module and the vision model can be applied independently, they can also be improved independently.

## 5. CONCLUSIONS

1. The three models tested in this paper are largely different, both in theoretical approach and in outcome. This means that the choice of a model has a large impact on sensor design decisions or trade-offs between sensors.
2. The following examples illustrate the difference in performance predictions:
  - a) When a 2 by 2 microscan is applied to a 'typical' under-sampled sensor, the TOD model predicts a large improvement (70% for high target contrasts), NVTherm predicts a minor improvement (about 10%), and TRM3 predicts a *decrease* in performance (minus 20%).
  - b) In the low contrast region, the  $\Delta T$  of the MTDP and TOD varies inversely proportional to the size of the test pattern, but the  $\Delta T$  of the MRTD-ID is constant.
  - c) The MTDP and MRTD-ID are different even in the well-sampled region where they both should be equivalent to the conventional MRTD.
3. Up to now, validation studies are limited and not comprehensive enough to determine which of the three models best predicts TA performance for real targets. At low contrasts, the MTDP and MRTD-ID have never been validated against the performance for real targets.
4. The differences between the predictions are so pronounced that we propose to perform a joint TA experiment with simulated sensors to decide which model best predicts TA performance. Obvious variables are sensor parameters such as those used in the present study, system noise, and target contrast.
5. The NVTherm model deals with under-sampling using a mathematical measure of the amount of non-linear image distortion. The relationship between this measure and TA performance is assessed experimentally in observer experiments, and this is accounted for in the performance predictions. The disadvantage is that it can only be applied to sensor configurations that do not deviate too much from well-known sensors for which identification data have been collected, and it is not directly applicable to new sensor concepts. Another disadvantage is that there exists no simple laboratory measure that corresponds to the predicted curve.
6. The NVTherm program gives predictions that are based on extrapolation and may therefore be inaccurate. These predictions may affect e.g. design choices. It is advised to determine the region for which the model has been validated and at least inform the user if operating outside this region.

7. The main advantage of the TRM3 approach is that the corresponding lab measure (MTDP) is similar to the MRTD and can be measured with the same test equipment. Modeling seems sound, and the correspondence between measured and calculated MTDP's is usually good.
8. The TRM3 approach introduces a strange artefact which makes its usefulness doubtful: for realistic undersampled sensors the *improvement* of sampling (e.g. microscan) can lead to a *decrease* in predicted performance. This is not expected for real targets (but this should be validated). The artefact can easily lead to wrong design decisions!
9. The TOD concept has the advantage that it is very general: unlike the two other approaches its application is not restricted to a certain family of sensor types. New concepts or application of local or global image enhancement, boost, microscan or an unusual detector shape can be implemented without changing the other elements of the model. Further, the TOD vision model is not restricted to certain types of test patterns (equilateral triangles in this case). In principle, the same model can be applied to a set of real targets.
10. The sensor module and the vision model of the TOD model can be applied independently, and can be improved independently. A possible application of the vision module is to replace the real observer in laboratory sensor performance measurements.
11. The TOD model is not yet ready for public release and the computations are (still) time-consuming because the software is not yet optimized for speed.

## REFERENCES

1. STANAG 4349.
2. Vollmerhausen, R., & Driggers, R.G (1999). NVTherm: next generation night vision model. Proc. IRIS Passive Sensors, 1, 121-134.
3. Bijl, P.& Valeton, J.M. (1998). TOD, the alternative to MRTD and MRC. Optical Engineering 37, 7, 1976 - 1983.
4. Wittenstein, W. (1999). Minimum temperature difference perceived – a new approach to assess undersampled thermal imagers. Optical Engineering 38, 5, 773 – 781.
5. Steven K. Moyer, Ronald G. Driggers, Richard Vollmerhausen, Keith Krapels (2001). Target Identification Performance as a Function of Spurious Response: Aliasing With Respect to the Half-Sample Rate. Proceedings SPIE 4372.
6. Devitt, N., Ronald G. Driggers, Richard Vollmerhausen, Steve Moyer, Keith Krapels & John O'Connor (2001). Target Recognition Performance as a Function of Sampling. Proceedings SPIE 4372.
7. Bijl, P.& Valeton, J.M. (1998). Validation of the new TOD method and ACQUIRE model predictions using observer performance data for ship targets. Optical Engineering 37, 7, 1984 - 1994.
8. Bijl, P., Valeton, J.M. & de Jong, A.N. (2000). TOD predicts target acquisition performance for staring and scanning thermal imagers, SPIE Proceeding Vol. 4030, 96-103.
9. Hogervorst, M.A., Bijl, P. & Valeton, J.M. (2001). Capturing the sampling effects: a TOD sensor performance model. SPIE Proceedings Vol. 4372, 62-73.
10. Night Vision Thermal Imaging Systems Performance Model NVTherm. Version 1.0. March 27, 2001.
11. Koenderink, J.J. and van Doorn, A.J. (1978). Visual detection of spatial contrast; Influence of location in the visual field, target extent and illuminance level. Biol. Cybern. 30, 157-167.