contributions to remote sensing: applications of thermal infrared

by

H. W. Brunsveld van Hulten (ed.)
P. Hoogeboom
A. F. G. Jacobs
C. Kraan
G. P. de Loor
and
L. Wartenaa

no. 34/1984
All correspondence should be addressed to

RIJKSWATERSTAAT
DIRECTIE WATERHUISHOUDING EN WATERBEWEGING
P.O.B. 20907
2500 EX THE HAGUE - THE NETHERLANDS

The views in this article are the authors' own.
RIJKSWATERSTAAT COMMUNICATIONS

CONTRIBUTIONS TO REMOTE SENSING: APPLICATIONS OF THERMAL INFRARED

H. W. Brunsveld van Hulten (ed.)
P. Hoogeboom
A. F. G. Jacobs
C. Kraan
G. P. de Loor
L. Wartena

The Hague 1984
Beechcraft-Queen Air survey aircraft of the Dutch National Aerospace Laboratory, used for all remote-sensing missions described. The pod under the fuselage houses an HBR-Singer Reconolax VI type infrared line scanner.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of symbols</td>
<td>8</td>
</tr>
<tr>
<td>List of abbreviations</td>
<td>10</td>
</tr>
<tr>
<td>Foreword</td>
<td>11</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>13</td>
</tr>
<tr>
<td>1.1 The need for new information systems</td>
<td>13</td>
</tr>
<tr>
<td>1.2 Remote sensing in the thermal infrared window</td>
<td>14</td>
</tr>
<tr>
<td>1.3 Applications investigated</td>
<td>16</td>
</tr>
<tr>
<td>2 Radiation temperature</td>
<td>18</td>
</tr>
<tr>
<td>2.1 Physical background</td>
<td>18</td>
</tr>
<tr>
<td>2.1.1 Planck’s law</td>
<td>18</td>
</tr>
<tr>
<td>2.1.2 Integration over a hemisphere</td>
<td>18</td>
</tr>
<tr>
<td>2.1.3 Integration over all wavelengths</td>
<td>19</td>
</tr>
<tr>
<td>2.1.4 Maximum-energy wavelength</td>
<td>20</td>
</tr>
<tr>
<td>2.1.5 Non-ideal radiators</td>
<td>20</td>
</tr>
<tr>
<td>2.2 Atmospheric effects</td>
<td>20</td>
</tr>
<tr>
<td>2.3 Temperature measurements</td>
<td>22</td>
</tr>
<tr>
<td>2.3.1 The radiation thermometer</td>
<td>22</td>
</tr>
<tr>
<td>2.3.2 Energy received by the radiation thermometer</td>
<td>22</td>
</tr>
<tr>
<td>2.3.3 Radiation temperature</td>
<td>23</td>
</tr>
<tr>
<td>2.3.4 True radiation temperature at the measuring level</td>
<td>24</td>
</tr>
<tr>
<td>2.3.5 Calculation of the temperature of the object</td>
<td>24</td>
</tr>
<tr>
<td>2.3.6 Relationship between the true and the measured temperature of the object</td>
<td>24</td>
</tr>
<tr>
<td>2.3.7 Correction for the atmospheric influence</td>
<td>25</td>
</tr>
<tr>
<td>2.4 Radiation temperature of water surfaces</td>
<td>26</td>
</tr>
</tbody>
</table>
2.4.1 Emission coefficient of water ........................................ 26
2.4.2 Skin temperature and bucket temperature .................. 29
2.4.3 Surface pollution of the sea ........................................ 32
2.4.4 Variations of the sea surface temperature .................. 32
2.5 Radiation temperature of the ground ......................... 33
2.5.1 Factors that exert an effect ..................................... 33
2.5.2 Structure of the vegetation cover ............................ 34
2.6 Characteristics of radiation thermometers and scanners .... 37
2.6.1 Infrared radiation thermometers ............................... 37
2.6.2 Infrared line scanners ........................................... 38
2.6.3 Infrared conical scanners ....................................... 40
2.7 Surface truth measurements ....................................... 41
2.7.1 Surface truth measurements on land .......................... 42
2.7.1.1 The field station ............................................. 42
2.7.1.2 Results ..................................................... 44
2.7.2 Surface truth measurements at sea ........................... 47
2.8 Image enhancement .................................................. 48

3 Applications .................................................................. 51

3.1 Hydrology ............................................................... 52
3.1.1 Seepage of salt water into agricultural areas .............. 52
3.1.2 Seepage of water due to high river discharge ............. 57
3.1.3 Seepage of water into sedimentary basins ................. 62
3.1.4 Seepage of water from waterways ....................... 64
3.2 Oceanography ........................................................... 67
3.2.1 Sea surface temperature ......................................... 67
3.2.2 Large-scale circulation .......................................... 73
3.2.3 Large-scale annual variations ................................. 78
3.3 Environment ............................................................. 83
3.3.1 Cooling-water discharges from power stations ........... 83
3.3.1.1 The Velsen power station ................................. 83
3.3.1.2 The Flevo power station ................................ 84
3.3.1.3 The Amer power station and the Moerdijk industrial area .... 86
3.4 Miscellaneous ............................................................ 92
3.4.1 Cooling water discharges from chemical plants ........... 92
3.4.2 Oil spill detection .................................................. 93
3.4.3 Navigation ............................................................ 99
3.4.4 Horizontal mixing ................................................ 101
3.4.5 Detection of ice ................................................... 102
Conclusions.................................................. 104

Appendix: The soil heat model.......................... 106

References................................................... 115
## List of symbols

Note: subscript $a$ relates to air  
subscript $s$ relates to soil  
subscript $w$ relates to water

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Units</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_a$</td>
<td>–</td>
<td>thermal diffusivity of air</td>
</tr>
<tr>
<td>$a_s$</td>
<td>–</td>
<td>thermal diffusivity of the soil</td>
</tr>
<tr>
<td>$c$</td>
<td>–</td>
<td>cloudiness ($0 \leq c \leq 1$)</td>
</tr>
<tr>
<td>$c_{p,a}$</td>
<td>J · kg$^{-1}$ · K$^{-1}$</td>
<td>specific heat of soil</td>
</tr>
<tr>
<td>$c_{p,s}$</td>
<td>J · kg$^{-1}$ · K$^{-1}$</td>
<td>specific heat of the soil</td>
</tr>
<tr>
<td>$C_D$</td>
<td>–</td>
<td>drag coefficient, bulk transfer coefficient for turbulence energy ($\approx 1.25 \times 10^{-3}$)</td>
</tr>
<tr>
<td>$C_L$</td>
<td>–</td>
<td>bulk transfer coefficient for the latent heat in air ($\approx 1.45 \times 10^{-3}$)</td>
</tr>
<tr>
<td>$C_H$</td>
<td>–</td>
<td>bulk transfer coefficient for the sensible heat in air ($\approx 1.45 \times 10^{-3}$)</td>
</tr>
<tr>
<td>$D$</td>
<td>m</td>
<td>damping depth</td>
</tr>
<tr>
<td>$E_v$</td>
<td>W · m$^{-2}$</td>
<td>irradiance integrated over all wavelengths</td>
</tr>
<tr>
<td>$E_{2,\lambda}$</td>
<td>W · m$^{-2}$ · m$^{-1}$</td>
<td>irradiance</td>
</tr>
<tr>
<td>$G_{\lambda}$</td>
<td>W · m$^{-2}$</td>
<td>scattered sky radiation measured in a horizontal plane</td>
</tr>
<tr>
<td>$h$</td>
<td>m</td>
<td>distance</td>
</tr>
<tr>
<td>$H$</td>
<td>m</td>
<td>distance between the sensor and the object</td>
</tr>
<tr>
<td>$k_s$</td>
<td>J · m$^{-1}$ · s$^{-1}$ · K$^{-1}$</td>
<td>thermal conductivity of the soil</td>
</tr>
<tr>
<td>$k_w$</td>
<td>J · m$^{-1}$ · s$^{-1}$ · K$^{-1}$</td>
<td>thermal conductivity of water</td>
</tr>
<tr>
<td>$L$</td>
<td>m</td>
<td>stability parameter (Monin-Obukhov length)</td>
</tr>
<tr>
<td>$L_{e,\lambda}$</td>
<td>W · m$^{-2}$ · sr$^{-1}$ · m$^{-1}$</td>
<td>radiance</td>
</tr>
<tr>
<td>$L_{ev}$</td>
<td>J · kg$^{-1}$</td>
<td>evaporation heat of water ($= 2.5 \times 10^{-6}$ J · kg$^{-1}$)</td>
</tr>
<tr>
<td>$m_w$</td>
<td>cm or g · cm$^{-2}$</td>
<td>amount of precipitable water</td>
</tr>
<tr>
<td>$e$</td>
<td>mbar</td>
<td>vapour pressure of water</td>
</tr>
<tr>
<td>$q$</td>
<td>g · kg$^{-1}$</td>
<td>specific humidity</td>
</tr>
<tr>
<td>$Q^*$</td>
<td>W · m$^{-2}$</td>
<td>net flux of radiant energy</td>
</tr>
<tr>
<td>$Q_{La}$</td>
<td>W · m$^{-2}$</td>
<td>latent-heat flux in air</td>
</tr>
<tr>
<td>$Q_{Ls}$</td>
<td>W · m$^{-2}$</td>
<td>latent-heat flux in soil</td>
</tr>
<tr>
<td>$Q_{Ha}$</td>
<td>W · m$^{-2}$</td>
<td>sensible-heat flux in air</td>
</tr>
<tr>
<td>$Q_{Hs}$</td>
<td>W · m$^{-2}$</td>
<td>sensible-heat flux in soil</td>
</tr>
</tbody>
</table>
\( Q_{\text{ir}} \) \( \text{W} \cdot \text{m}^{-2} \) net long-wave radiation flux

\( r_\lambda \) - reflection coefficient (reflectance) of the object

\( T \) \( \text{K}, \, ^\circ \text{C} \) temperature

\( T_0 \) \( \text{K}, \, ^\circ \text{C} \) temperature of the surface layer of the object (‘skin temperature’)

\( T_a \) \( \text{K}, \, ^\circ \text{C} \) temperature of the air between the object and the sensor

\( T_c \) \( \text{K} \) cloud-base temperature

\( T_{\text{cal}} \) \( ^\circ \text{C} \) temperature of the immediate surroundings during calibration

\( T_e \) \( \text{K}, \, ^\circ \text{C} \) temperature of the objects’ surrounding

\( T_m \) \( ^\circ \text{C} \) temperature measured by the instrument used

\( T_{\text{meas}} \) \( ^\circ \text{C} \) temperature of the immediate surroundings during measurement

\( T_r \) \( \text{K} \) radiation temperature

\( T_r(H) \) \( \text{K} \) true radiation temperature at the measuring level

\( T_s \) \( \text{K}, \, ^\circ \text{C} \) soil temperature

\( T_w \) \( ^\circ \text{C} \) bulk or ‘bucket’ temperature

\( u_\text{a} \) \( \text{m} \cdot \text{s}^{-1} \) friction velocity

\( u_{10} \) \( \text{m} \cdot \text{s}^{-1} \) mean wind velocity 10 m above the ground surface

\( z_{\text{max}} \) \( \text{m} \) height of the air layer under consideration where there is no longer any exchange with higher air layers

\( z_0 \) \( \text{m} \) reference surface layer

\( x_\lambda \) - absorption coefficient (absorptivity)

\( \Delta T \) \( \text{K} \) temperature difference

\( e_\lambda \) - emission coefficient (emissivity) of the object

\( \vartheta \) rad, deg angle between observation direction and the object’s surface normal

\( \kappa \) - Kármán constant (\( \sim 0 \cdot 4 \))

\( \lambda \) \( \mu \text{m}, \text{m} \) wavelength of the radiation

\( \lambda_{\text{max}} \) \( \mu \text{m}, \text{m} \) wavelength of maximum energy emission

\( \rho_a \) \( \text{kg} \cdot \text{m}^{-3} \) density of the air

\( \rho_s \) \( \text{kg} \cdot \text{m}^{-3} \) density of the soil

\( \sigma \) \( \text{W} \cdot \text{m}^{-2} \cdot \text{K}^{-4} \) Stefan-Boltzmann constant

\( = 5.67 \times 10^{-8} \text{W} \cdot \text{m}^{-2} \cdot \text{K}^{-4} \)

\( \tau_\lambda \) - transmission coefficient (transmissivity) of the object

\( \tau_\lambda(h) \) - transmission coefficient (transmissivity) of air over a distance \( h \)

\( \Phi_\lambda \) - spectral sensitivity of the sensor

\( \omega, \Omega \) \( \text{sr} \) solid angle

\( \omega_1 \) \( \text{rad} \cdot \text{s}^{-1} \) angular frequency of the first harmonic of the daily temperature cycle at the surface

\( v_w \) \( \text{m}^2\text{s}^{-1} \) kinematic viscosity of water
### List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSRAD</td>
<td>absolute radiation thermometer</td>
</tr>
<tr>
<td>HW</td>
<td>high water</td>
</tr>
<tr>
<td>IR</td>
<td>infrared</td>
</tr>
<tr>
<td>IRCS</td>
<td>infrared conical scanner</td>
</tr>
<tr>
<td>IRLS</td>
<td>infrared line scanner</td>
</tr>
<tr>
<td>IRT</td>
<td>infrared radiation thermometer</td>
</tr>
<tr>
<td>KNMI</td>
<td>Royal Netherlands Meteorological Institute</td>
</tr>
<tr>
<td>LV</td>
<td>lightvessel</td>
</tr>
<tr>
<td>LW</td>
<td>low water</td>
</tr>
<tr>
<td>NAP</td>
<td>Normaal Amsterdams Peil ('Normal Amsterdam Water Level') – a basic ordnance datum used in the Netherlands</td>
</tr>
<tr>
<td>NETRAD</td>
<td>net radiation thermometer</td>
</tr>
<tr>
<td>NLR</td>
<td>National Aerospace Laboratory NLR</td>
</tr>
<tr>
<td>RLD</td>
<td>National Aviation Service</td>
</tr>
<tr>
<td>RT</td>
<td>radiation thermometer</td>
</tr>
<tr>
<td>SST</td>
<td>sea surface temperature</td>
</tr>
<tr>
<td>TNO</td>
<td>Netherlands Organization for Applied Scientific Research</td>
</tr>
<tr>
<td>TSSL</td>
<td>temperature of the sea surface layer</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
</tbody>
</table>
Foreword

Being a major potential user of remote sensing techniques in the Netherlands, the Rijkswaterstaat has since the early stages (1968) played an important part in the studies on the applications of these techniques in fields relevant to its activities. Since this is very much a multidisciplinary subject various institutes have cooperated in the work, in particular the Physics Laboratory TNO, which has had years of experience in this field.

Nearly all the aerial observations by the National Aerospace Laboratory NLR were made from its Beechcraft-Queen Air type aircraft (it now uses a Swearingen-Metro II aircraft). One experiment was carried out by the National Aviation Service, using a Dakota over a period of a year.

This publication summarizes the results of remote sensing in the thermal infrared region carried out in the period 1971–78. The report is the work of the following persons:

H. W. Brunsveld van Hulten (Ed.) Department of Water Management and Hydraulic Research, Rijkswaterstaat, The Hague, now Pan-Consult BV, Delft

Dr. G. P. de Loor and P. Hoogeboom Physics Laboratory TNO, The Hague

Prof. Dr. L. Wartena and Dr. A. F. G. Jacobs Research Department, The Netherlands Development and Reclamation Society, Arnhem, both now at the Department of Physics and Meteorology, University of Agriculture, Wageningen

C. Kraan Royal Netherlands Meteorological Institute (KNMI), De Bilt
1 Introduction

1.1 The need for new information systems

The Rijkswaterstaat is responsible for defense against the forces of water, control of the water economy, and maintenance of the water quality. Its work covers the area within the Dutch borders and on the Dutch part of the Continental Shelf in the North Sea, and requires a great deal of qualitative and quantitative information. Such information can be obtained in different ways, depending on the purpose it is to serve. Complicated information-gathering systems are needed to obtain data on complicated phenomena such as hydraulics, traffic, inundation, and the pollution of surface waters, all of which are aspects that the State desires, and has a duty, to control. The means of observation are generally chosen to suit the nature of the phenomena to be observed, and these – notably natural phenomena – are dynamic and mostly very complicated.

As society becomes more and more complex, the Rijkswaterstaat acquires an increasing number of functions, and people become ever more demanding about the way these are to be carried out. One must therefore similarly become more and more demanding about the means of observation used to obtain the necessary information, so that the observations become quicker, more accurate, more detailed, preferably cover large areas, and are independent of the weather.

For this reason, the Rijkswaterstaat started as early as 1968 to study the possibilities of remote sensing in a number of relevant fields. The present publication deals with the use of thermal infrared radiation, and particularly, of thermal mapping.

Although the literature mentions many applications of thermal mapping that have shown varying degrees of success, only some large-scale operational use of this method has yet been reported. When research on remote sensing first started in the Netherlands, the criticism was voiced that a great deal of image processing was being done with complicated computer systems, but that not enough knowledge had been gathered about the object-sensor relationship, which had hardly been investigated at all. For this reason, and because of financial and manpower constraints, we have now paid much less attention to image processing, and in fact this topic will be treated here only as far as necessary.

The work described in this report covered the previously neglected aspects mentioned and yielded basic information both about certain processes that take place in and on
objects forming the targets of remote sensing and about the ways the signals are changed by the atmosphere and the sensor system before they form an image. The value of these results is that, if the object-sensor relationship and the transfer function of the sensor system are known, the system can be optimized for a given application so that it eventually can become a (near) real-time and dedicated system at a reasonable cost. If a general-purpose system is built and used without a sufficient knowledge of the sensor-object relationship, it becomes necessary to collect a surfeit of data, which can be processed only with the aid of a very large computer system. Besides, the results then obtained are often no better than ones that can be secured by the ordinary interpretation of images. In the past, expensive remote sensing techniques often failed to find acceptance as new measuring and surveying techniques in some well-established research fields because of the bottleneck caused by the data processing and because of the high computer costs in comparison with the amount of information finally obtained. Thus, it is only recently that thermal infrared sensing has gained recognition in the earth sciences.

1.2 Remote sensing in the thermal infrared window

Thermal radiation consists of electromagnetic waves of a certain length (see Fig. 1). The intensity and the direction of this radiation, which is emitted by everything in nature, are sensed and registered by infrared detectors. The information thus obtained (the

![Figure 1. The electromagnetic spectrum (Reeves 1975).](image-url)
Radiation temperature is used to determine the surface temperature of the ground and water and of any objects present on them, such as vegetation and ships. In most applications one is interested not only in the (absolute) surface temperature and the horizontal temperature differences, but also in the properties of the surface that which together with many other parameters determine its radiation temperature.

The surface temperature is determined by the thermal state of the object, with its variations, and by the thermal processes in the object's surroundings, with which it exchanges energy. In the boundary layers between the ground and the air, and between water and air, the nature and composition of the interface or the interfacial layer are the most decisive factors in determining the radiation temperature. A good understanding of these heat balances and the influential characteristics of the boundary layer is essential for the determination of the surface temperature from the radiation temperature.

The mean temperatures of the ground, oceans, and marginal seas follow annual cycles that are out of phase with the annual cycle of the mean atmospheric temperature. A great variety of phenomena involving energy changes occur in the atmosphere, depending on the geographical latitude. These processes, which cause fairly rapid temperature fluctuations, include insolation, cloud formation, cooling, and rapid horizontal and vertical movements of air masses at different temperatures. Although the sea moderates the temperature fluctuations of the adjoining land through the air, the energy fluctuations and the relationships between them exert a strong influence on the radiation temperature. The changes occur at intervals of a few days in the case of the atmosphere, at intervals of a few weeks in the case of the ground, vegetation, snow cover, and humidity, and at intervals of some months or years in the surface layers of oceans, increasing to millennia at great depths. It is brought into mind that there exists a relationship between temporal and spatial scales. These factors, which along with some others ultimately determine the radiation temperature of the Earth's surface, are interlinked in an extremely complicated manner via heat, momentum, and water exchanges in all their possible manifestations.

As thermal radiation traverses the atmosphere some of its energy is lost through scattering and absorption, but at the same time the air itself makes a contribution to the signal on account of its thermal state. Owing to evaporation processes, after rain and fog this contribution is so dominant that no thermal mapping can then be carried out, effectively. However, night-time conditions are of no restriction to thermal mapping. Owing to the complexity of atmospheric processes, the description of their effect on the recorded radiation involves a large element of uncertainty. In this connection, the thickness of air layer through which the observation is made is of importance.

These comments indicate that we need a good understanding of the processes involved
if we are to decide in practice at what time of the day and under what weather conditions we can collect information with a maximum chance of obtaining significant temperature differences between different terrain elements, or of being able to observe the phenomena for a sufficiently long time.

When discussing the applications we shall examine the nature of the information obtained and show that the thermal infrared method cannot yet be regarded as an independent survey technique. For the time being it provides additional information, which – when taken in conjunction with the surface truth – can lead to a more complete picture of the phenomena in question.

1.3 Applications investigated

Thermal mapping has been found useful in observing the temperature distributions over large areas; this makes it possible to observe not only objects that stand out against their surroundings or background by their radiation temperature, but also morphological structures and vegetation patterns. When the temperature contrasts are small, one must first establish when they are likely to be most pronounced. This is done e.g. for investigating hydrological phenomena such as seepage under and through dikes and the seepage of water through covered river beds in sedimentary basins.

The surface structures are connected with deeper layers of complicated structures, which rule percolation and seepage. The morphological and geological surface structures therefore give rise to differences in the radiation temperature, and thermal mapping can thus yield indirect information about seepage.

Thermal mapping can also prove useful in statistical studies of shipping, mainly in busy harbour entrances and shipping lanes, by supplying information on the types of the vessels, their speeds, and their courses, from observations made at successive times.

The temperature can be used as a tracer, e.g. to follow the movements of water in coastal areas and in tidal rivers. This possibility has been utilized to monitor the spread of warmer water when industrial cooling water flows into rivers, canals, the IJsselmeer, the Hollands Diep, etc. The method has been used to investigate the movements of water in the coastal area near the Hook of Holland and the island of Texel, and to study the large-scale temperature distribution along the coast of the Province of Holland over a period of one year.

Since oil and water have different emission coefficients (emissivities), it is possible to detect accidental or deliberate oil spills at sea, and a number of experiments have in fact been carried out in this connection.
Thermal mapping in itself is not sufficient to detect a combination of spilled oil and cooling water, one that is often found in ports with chemical plants; this calls for the combined application of a number of remote sensing techniques.
2 Radiation temperature

2.1 Physical background

2.1.1 Planck’s law

To understand the concept of radiation temperature we must first consider an ideal body called a black body.

A black body is one that absorb all radiation incident on it, irrespective of the wavelength. The radiance $L_{e,\lambda}(T)$ of a black body at temperature $T$ is the amount of radiation energy of wavelength $\lambda$ that it emits in unit time, into a unit solid angle, in unit wavelength interval, and from a unit surface area normal to the direction of observation.

$$L_{e,\lambda}(T) = \frac{C_1 \lambda^{-5}}{\exp\left(\frac{C_2}{\lambda T}\right) - 1} \quad (1)$$

where:

\begin{align*}
C_1 &= 1 \cdot 1909 \cdot 10^{-16} \text{ W} \cdot \text{m}^2 \\
C_2 &= 1 \cdot 4388 \cdot 10^{-2} \text{ K} \cdot \text{m}
\end{align*}

Please refer list of symbols.

2.1.2 Integration over a hemisphere

In the case of uniform radiation, i.e. when $L_{e,\lambda}(\theta) = L_{e,\lambda}(0^\circ) \cos \theta$, the irradiance $E_{e,\lambda}(T)$ of a black body, integrated over a hemisphere $\Omega$, is given by:

$$E_{e,\lambda}(T) = \int_{\Omega} L_{e,\lambda}(T) \cdot \cos \theta \, d\Omega = \pi L_{e,\lambda}(T) \quad (2)$$

where $\theta$ is the angle between the normal and the direction of observation. Fig. 2 gives $E_{e,\lambda}(T)$ for $T = 308 \text{ K}$ and $T = 271 \text{ K}$. 
2.1.3 Integration over all wavelengths

Integration of the irradiance over all wavelengths gives the well-known Stefan-Boltzmann equation describing the relationship between the temperature of a black body and the total radiation energy that it emits into a hemisphere:

$$E_e(T) = \int_0^{\infty} E_{e,\lambda}(T) d\lambda = \sigma T^4$$

(3)

where $\sigma$ is the Stefan-Boltzmann constant.

The basic equation for all radiation-temperature measurements techniques is derived from (3) by confining the integration over a narrow band ($\lambda_1, \lambda_2$). The band-width is chosen according to theoretical considerations and practical limitations of the sensor applied, thus

$$E'_e(T) = \int_{\lambda_1}^{\lambda_2} E_{e,\lambda}(T) d\lambda$$

(3A)
2.1.4 Maximum-energy wavelength

The wavelength at which maximum energy is emitted and which varies with the temperature of the object \( T \), is found by differentiating Eq. (1):

\[
\lambda_{\text{max}}(T) = 2897 \cdot 8T^{-1}
\]  \( \text{(4)} \)

For a black body at room temperature \( (T \approx 300 \text{ K}) \), the value of \( \lambda_{\text{max}} \) is about 10 \( \mu \text{m} \). This corresponds to the mid-point of one of the atmospheric ‘windows’ (see Section 2.2.). As Fig. 2 shows, the value of \( \lambda_{\text{max}} \) is about 9.4 \( \mu \text{m} \) at 35°C and about 10.7 \( \mu \text{m} \) at -2°C; the atmospheric window in question extends from 8 to 14 \( \mu \text{m} \).

2.1.5 Non-ideal radiators

There are no ideal black bodies in nature. All real bodies emit less energy than the ideal body considered in Section 2.1.3. Certain real bodies, for example the atmosphere, are to some extent also transparent to thermal radiation. The behaviour of a body towards radiation is defined by three scalar quantities, namely the reflection coefficient \( r \), the transmission coefficient \( t \), and the absorption coefficient \( a \). All three coefficients depend on the wavelength, which is indicated by a subscript \( \lambda \). By definition these coefficients represent the fractions of the original energy that are reflected, transmitted, and absorbed by the body in a certain wavelength band, and according to conversation of energy there holds:

\[
r_{\lambda} + t_{\lambda} + a_{\lambda} = 1
\]  \( \text{(5)} \)

The emission coefficient \( \varepsilon_{\lambda} \) is always equal to the absorption coefficient \( a_{\lambda} \). If \( t_{\lambda} = 0 \), the reflection coefficient is \( r_{\lambda} = 1 - a_{\lambda} = 1 - \varepsilon_{\lambda} \). For a black body \( r_{\lambda} = t_{\lambda} = 0 \), and so \( a_{\lambda} = \varepsilon_{\lambda} = 1 \).

2.2 Atmospheric effects

The atmosphere is opaque to the electromagnetic radiation coming from the sun, except in four regions of the spectrum: the near ultra violet, the visible, the near-infrared, and the far-infrared, (see Sect. 1.2, Fig. 1). The energy reaching the surface of the Earth is partly reflected and partly absorbed. The incoming solar radiation (insolation) heats the ground, and the latter in turn emits long wave (thermal) radiation. At normal surface temperatures of 300 K this emission has, as we saw, a maximum at a wavelength of about 10 \( \mu \text{m} \), which lies in the ‘atmospheric window’ extending between 8 and 14 \( \mu \text{m} \) (see Sect. 2.1.2, Fig. 2). In this spectral region the
emitted thermal radiation dominates over the reflected solar radiation (see Sect. 2.5.1, Fig. 9).

The temperature of terrestrial surfaces (land or water) is found from the amount of emitted thermal radiation measured by the sensor, taking into account the changes in the signals as they traverse the atmosphere. Apart from scattering, the atmospheric effects can be explained as follows. The atmosphere is composed of 78% of nitrogen (N\(_2\)), 21% of oxygen (O\(_2\)), 0.02-0.04% of carbon dioxide (CO\(_2\)), some water vapour (H\(_2\)O), and trace elements such as ozone (O\(_3\)) and aerosols. Electromagnetic radiation with a wavelength longer than that of X-rays interacts both with atoms and with molecules (Shanda, 1976; Gjessing, 1978). As a result, energy is variously absorbed and released, giving rise to absorption and emission spectra. The electrons forming part of an atom or molecule can jump to higher energy levels when they absorb a quantum of energy, but the probability of this happening is small, due to the low energy levels in the IR wavelength. It is much more likely that the whole molecule will be raised to a higher energy state. This happens in particular with triatomic molecules like CO\(_2\), H\(_2\)O and O\(_3\), which have both rotational and vibrational states. These states are quantized, and less energy is needed for transitions in the rotational range than for transitions in the vibrational range, so that the resulting absorption spectra can be readily differentiated. In general, electron transitions result from interactions with UV and visible light, while changes in the vibrational states are connected with the near infrared and changes in the rotational states are connected with the far thermal infrared and the microwave region (see Fig. 1). The spectra resulting from rotational transitions have broad bands, so that such spectra can be used in the thermal infrared region.

Since the atmosphere consists of a number of different gases, the spectrum exhibits various absorption bands. Fig. 1 (Sect. 1.2), shows a representative spectrum of the atmosphere, from which it can be seen that the atmosphere is transparent to thermal radiation only in the regions of 3-5 and 8-14 \(\mu\)m. The first of these windows lies on the steep slope at the left-hand side of the Planck radiation curve. Some measurements are carried out in this band, since the radiation differences are larger though the absolute energy levels are lower. Detectors are also more sensitive in this region, and the emission coefficients show greater differences than in the 8-14 \(\mu\)m band. Both windows are used at the same time for studying the emission coefficients. In general, however, the measurements are carried out in the 8-14 \(\mu\)m band, because a) the absolute energy level is higher here, b) the emission coefficients do not exhibit excessive differences, and c) most detectors have been adapted for work in this region, and d) less solar radiation is reflected. Absorption occurs on both sides of the band: on the short-wave side (\(\lambda < 8 \mu m\)), it is due to water molecules, while on the long-wave side (\(\lambda > 13 \mu m\)) it is due to both carbon dioxide and water molecules; in the range of 8-13 \(\mu\)m ozone can exert some effect. The band-width used should be as large as possible, to ensure that the energy levels to be measured are not too low, i.e. to guarantee a favourable signal-to-
noise ratio. To minimize the generally stochastic absorption effects, the side-bands mentioned above must be eliminated by the use of optical filters. This is why different filter band-widths are used, covering regions such as 8–14 \( \mu \text{m} \) and 9.5–11.5 \( \mu \text{m} \) (see Section 2.6).

The concentration of aerosols causing scattering of the IR radiation may vary strongly and is hard to establish. Some practical approaches to corrections for aerosol scattering, always combined with the atmospheric absorption, are discussed in Section 2.3.7.

2.3 Temperature measurements

2.3.1 The radiation thermometer

The radiant energy emitted by an object can be measured with a radiation thermometer or radiometer. This instrument measures over a narrow solid angle \( \omega \) (corresponding to the aperture of the instrument) and within a wavelength interval \((\lambda_1, \lambda_2)\) determined by the filter used. The radiometer gives an output signal that varies with the amount of energy received and which is related to the temperature of the emitting object as shown by Eq. (3A). The instrument is generally calibrated against a black body in the laboratory. Some IR radiometers and scanners are described in more detail in Section 2.6.

2.3.2 Energy received by the radiation thermometer

The energy received by the radiometer comes from three different sources: from the object itself, from the object’s surroundings and from the air traversed.

The air between the radiometer and a distant object both adds to and subtracts from the radiation energy emitted by the object (see Fig. 3). Furthermore, some radiation from the object’s surroundings reaches the instrument after reflection from the object itself. The reflection of radiation coming directly from the sun usually can be neglected in the 8–14 \( \mu \text{m} \) window (see Sect. 2.5.1, Fig. 9). The energy received is thus the sum of the above three components:

\[
E_{e,\text{received}}(H) = \omega \int_{\lambda_1}^{\lambda_2} \varepsilon_{\lambda} \cdot \Phi_{\lambda} \cdot \tau_{\lambda}(H) \cdot L_{e,\lambda}(T_o) \cdot d\lambda \\
+ \omega \int_{\lambda_1}^{\lambda_2} (1 - \varepsilon_{\lambda}) \cdot \Phi_{\lambda} \cdot \tau_{\lambda}(H) \cdot \frac{G_{\lambda}}{\pi} \cdot d\lambda \\
+ \omega \int_{\lambda_1}^{\lambda_2} \Phi_{\lambda} \cdot L_{e,\lambda}(T_a(h)) \cdot \frac{\partial \tau_{\lambda}(h)}{\partial h} \cdot dh \cdot d\lambda
\]

(6)
It should be borne in mind that the factor $\partial \tau_x(h)/\partial h$ is negative.

![Diagram](image)

Figure 3. Contributing factors to the incoming radiation measured by a radiometer.

2.3.3 Radiation temperature

The radiometer interprets the incident radiant energy as coming from a black body at a radiation temperature $T_r$:

$$E_{e,\text{received}}(T_r) = \int_{\lambda_1}^{\lambda_2} \Phi_\lambda \cdot L_{e,\lambda}(T_r) \cdot d\lambda$$  \hspace{1cm} (7)

To a first approximation it can be assumed that the emission coefficient $\varepsilon$ and the transmission coefficient $\tau$ are constant, as they are average or effective values over the interval $\lambda_1, \lambda_2$. We also assume for the scattered sky radiation that $G_\lambda = \pi L_{e,\lambda}(T_e)$. Integration of Eq. (6) and substitution of the terms involved in Eqs. (3) and (7) then give:

$$T_r^4(H) = \varepsilon \cdot \tau(H) \cdot T_o^4 + (1 - \varepsilon) \cdot \tau(H) \cdot T_e^4$$

$$+ (1 - \tau(H)) \cdot T_a^4(H)$$  \hspace{1cm} (8)

It follows, therefore, that the atmospheric influence and the reflection from the object must be subtracted from the radiometer reading to obtain the object's temperature. This will be discussed in detail in Sections 2.3.7 and 2.4.1.
2.3.4 True radiation temperature at the measuring level

The measured temperature $T_m$ must first be corrected on the basis of the laboratory calibration against a black body. A further correction is needed if the ambient temperature during measurement ($T_{\text{meas}}$) is different from that during calibration ($T_{\text{cal}}$). The true radiation temperature is therefore obtained as follows:

\[
T_r(H) = T_m + \Delta(\text{cal}) + f \cdot (T_{\text{cal}} - T_{\text{meas}}) \tag{9}
\]

The factor $f$ must be determined experimentally, e.g. in a controlled climate room. For a Barnes PRT-5 radiometer with a 9.5–11.5 $\mu$m window, $f = 0.1$. The third term on the right-hand side of Eq. (9) is included to correct for the effect of the ambient temperature on the chopper blades rotating in front of the radiometer filter (‘shutter correction’), since this causes a deviation in the output signal (Kraan, 1977). This correction is not needed when the instrument is calibrated in the field during the measurements.

2.3.5 Calculation of the temperature of the object

The object temperature is found by combining Eqs. (6) and (7). However, the integration is not easy, and there are several ways to approximate the solution. In Tien’s useful method (1974), Eq. (8) is expanded into a first-order Taylor series, giving:

\[
T_o = T_r(H) + \frac{1 - \tilde{\varepsilon}}{\tilde{\varepsilon}}(T_r(H) - T_e) + \frac{1 - \tau}{\tilde{\varepsilon} \cdot \tau}(T_r(H) - T_d(H)) \tag{10}
\]

where the mean atmospheric transmission coefficient $\tilde{\tau}$ is determined by the amount of precipitable water in the atmosphere (see Section 2.3.7). The effective emission coefficient for clean water and for a filter band-width of 9.5–11.5 $\mu$m is $\tilde{\varepsilon} = 0.99$ (see Sect. 2.4). This quantity for soil shows a large scatter and is hard to establish (see Section 2.5.1).

2.3.6 Relationship between the true and the measured temperature of the object

Most objects are only slightly transparent to thermal radiation and the radiation emitted by them originates in the surface layer. This surface temperature $T_o$, obtained after correction, is generally different from the true internal temperature of the object, which in case of water is called the bulk or bucket temperature. If the surface layer is very thin ($10^{-2}$ m or less), the temperature measured is called the skin temperature. For such objects the radiometrically measured temperature cannot be checked by measuring the body temperature by conventional devices on location, giving what is called the
'surface truth' or the 'sea truth' (see Sects 2.3.2, Fig. 3 and 2.7.2, Fig. 19). The difference between the skin temperature and the bulk temperature depends to a great extent on the object's physical properties, a topic that will be discussed for water in Section 2.4 and for soil in Section 2.5.

2.3.7 Correction for the atmospheric influence

The two terms for atmospheric influences and reflected background radiation are given in Eq. (6). Atmospheric scattering and multiple reflections are ignored. The atmospheric influence often cannot be determined directly because of its complicated nature, even if the CO$_2$ content is assumed to be constant (which can be done if the measurement takes only a few hours) and only the variations in the water vapour concentration are taken into account. In practice the atmospheric influence is therefore mostly determined directly, by carrying out simultaneous measurements from an aeroplane or satellite and directly next to the object, preferably with the same type of infrared radiation thermometer (see Section 2.7). The difference between these two measurements gives a correction factor that can be extrapolated over the whole covered area.

The accuracy of the correction is comparable to the absolute accuracy of the infrared radiation thermometer. Using experimental data, Lorenz (1971a, 1971b, 1973) determined the relationship between the atmospheric correction and the radiation temperature $T_r(H)$ and the air temperature $T_a(H)$ at various levels $H$. In this he assumed that the atmosphere is not stratified, the relative humidity is constant, and the temperature gradient is also constant. The literature also contains other approximations, e.g. that given by Saunders (1970).

The following simple analytical approach is suggested. The mean atmospheric transmission coefficient $\bar{\tau}$ can be correlated with the mean amount of precipitable water $\bar{m}_w$ as follows:

$$\bar{\tau} = \exp(-0.18\bar{m}_w)$$

This holds for instruments with a band-width of 9.5–11.5 μm (Platt, 1972). At sea level and for $H < 300$ m, we also have:

$$\bar{m}_w = 1.25 \cdot 10^{-4} \cdot \bar{\varrho} \cdot H$$

where $\bar{\varrho}$ is the specific humidity.

For $H = 150$ m and for a filter band-width of 9.5–11.5 μm, $\bar{\tau} \approx 1 - 0.003\bar{\varrho}$. For an altitude of 300 m a 10% correction is necessary because of aerosols.

When the band-width is 8–14 μm, the above correction factors obtained with a filter
band-width of 9.5–11.5 μm must be multiplied by about 1.5.
Some authors combine atmospheric influences with the correction for the reflection (see Sect. 2.4.1).

2.4 Radiation temperature of water surfaces

2.4.1 Emission coefficient of water

Since water is almost entirely opaque to radiation with a wavelength of about 10 μm, the relation \( \varepsilon = 1 - r \) is valid. According to Lorenz (1973), 90% of the radiation is absorbed in the first 10–50 μm of water. The emission coefficient varies with the wavelength (see Fig. 4) and with the angle of observation ('elevation') in the range 30–90° (see Fig. 5).

![Figure 4. The emission coefficient of water as a function of wavelength.](image)

For clear freshwater the mean emission coefficient \( \bar{\varepsilon} \) measured vertically downward is about 0.99 and about 0.98 for \( \lambda = 9.5–11.5 \) μm and \( \lambda = 8–14 \) μm, respectively (Robinson and Davies, 1972; see Fig. 5). The influence of wave motion and direct reflected solar radiation is probably negligible (see Sect. 2.5.1, Fig. 9). The reflection coefficient of seawater is the same as that of freshwater when measured perpendicularly, but it is 3% or 1% higher at 60°, depending on whether a band-width of 9.5–11.5 μm or one of 8–14 μm is used.

The dependence of \( T_r(H) \) on the reflection coefficient can only be reduced by using horizontally polarized radiation at a Brewster angle of 54° (Lecomte et al., 1973).
The reflection coefficients of natural bodies of water are often very sensitive to the presence of even monomolecular surface layers (see Section 2.4.3). They are generally 0.02 lower in the presence of a thin oil layer on the water.

To correct for the reflected sky radiation we distinguish between a clear and an overcast sky. Eq. (8) applies to both, while Eq. (10) applies only to the latter. In the case of uniform, low, and thick clouds, the cloud base can be regarded as a black body, so that in Eq. (6):

$$\frac{1}{\pi} G_{\lambda} = L_{c,\lambda} (T_c)$$

where \(T_c\) is the radiation temperature of the cloud base. The temperature of the surroundings \(T_e\) is thus replaced by \(T_c\) in Eq. (10).

For a partially overcast sky we derive a weighted mean correction factor from the correction factors that apply to a clear and to a fully overcast sky.

The corrections for various types of cloud in the case of IRT's with a band-width of 8–15 \(\mu\)m have been calculated by Saunders (1970) (see Table 1). Under a broken cumulus cloud cover over the sea, the correction ('noise') in the local water temperature amounted to \(\pm 0.5\) K.
Frank (1964) approached this problem in the same way as did Fuchs and Tanner (1966): he made the same assumptions for $\varepsilon$ and $\tau$ and considered the clouds to behave as black bodies. He thus obtained Eq. (8) with $T_e = T_c$.

Table 1  Correction of the temperature for reflection from a water surface under various types of clouds (Saunders 1970).

<table>
<thead>
<tr>
<th>Cloud type</th>
<th>Cloud height [km]</th>
<th>Correction $T_o - T_r(0)(^\circ C)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clear</td>
<td>-</td>
<td>0.5...0.7</td>
</tr>
<tr>
<td>Dense Cirrostratus, overcast</td>
<td>8</td>
<td>0.4...0.55</td>
</tr>
<tr>
<td>Altocumulus or Altostratus, overcast</td>
<td>6</td>
<td>0.25...0.4</td>
</tr>
<tr>
<td>Stratus or Stratocumulus, overcast</td>
<td>3</td>
<td>0.2</td>
</tr>
<tr>
<td>Stratus or Stratocumulus, overcast</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>Stratus or Stratocumulus, overcast</td>
<td>1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Some authors use corrections that take account of both the atmospheric influence and the reflection together.

Figure 6. Diagrams for atmospheric correction, aerosols and ozone not included. Band-width 8-14 $\mu$m. (Shaw and Irbe 1972).
Shaw and Irbe (1972) constructed some correction diagrams (Fig. 6) by comparing the theoretical relationship in Eq. (6) against the experimental results without aerosols and the ozone effect. These diagrams show the variations in $[T_o - T_s(H)]$ in dependence of $[(T_s(h) - T_s(H)]$ and the total water-vapour content of the air layer in the case of clear and overcast skies using an instrument with a band-width of 8–14 μm (T_s(h) is the mean air temperature at a vertical distance h). The ‘60° test’ experimental method of Saunders (1967a) is also used to determine jointly the atmospheric influence and the reflection from airborne observations. Here use is made of two preferably identical infrared radiation thermometers, one pointed vertically down and the other forward or backward at an angle of 60°. The atmospheric path length of the signal for the second radiometer is twice as long as that for the first, so that the atmospheric influence is also roughly twice as large as for the first instrument. Saunders has found that the reflection coefficient of water is exactly twice as large when measured at an angle of 60° as it is when measured vertically downward in the case of a filter band-width of 8.25–12.35 μm (see Fig. 4). The difference between the two measurements for the same geographical point gives the combined correction. The atmosphere is assumed to be homogeneous. It should be added that the data on which Fig. 4 is based are probably not very reliable, and other authors give values of 2–4.5 for the ratio between the values of the reflection coefficient r at 60° and 0°.

This 60° correction method has been used in the Netherlands when measuring the circulation of sea water near the island of Texel and at the Hook of Holland (see Section 3.2.3). It provides correction for the atmospheric influence and the reflection along the whole flight line. It is suitable for observations over the sea, particularly when taking measurements at different altitudes over an area with an approximately uniform water surface temperature.

The method is sometimes used with a single infrared radiation thermometer, which is then alternately set vertically down and at 60°. It is advantageous if the temperature is uniform, but this is not essential. If the temperature is not uniform, the accuracy with which the measuring point can be fixed geographically also determines the accuracy of the measurements. The results of the experiments described in Section 3.2 were still insufficiently accurate. If the platform parameters (time and place) and the radiometric data are recorded digitally, the accuracy is expected to improve.

2.4.2 Skin temperature and bucket temperature

From Sections 2.3.6 and 2.4.1 it follows that the radiation temperature of a body of water is represented by that of an extremely thin layer on its surface, and because of the very small heat capacity and its direct contact with the atmosphere it can differ from the temperature of the deeper layers, measured with conventional instruments.
This surface layer, or skin layer, also differs from the bulk of the water in its characteristics, since laminar flow and viscosity effects predominate in it, in contrast to turbulent transport deeper down (see Fig. 7). In the case of the sea we thus have the sea surface temperature (SST) $T_o$ as the skin temperature, and the temperature of the sea surface layer (TSSL) $T_w$ as the bulk or bucket temperature.

![Figure 7. The energy budget of the insolation over the sea surface.](image)

As a result of evaporation and the emission of long-wave radiation, $T_o$ is generally about 0.5 K lower than $T_w$. The skin layer is fairly stable and re-forms quickly (in a time of the order of 10 s) following a disturbance (Ewing and McAlister, 1960). This means that, in contrast to the wind, the sea state has only a small effect on the skin temperature $T_o$.

Since we do not know the exact nature of the physical processes that take place at the air-sea interface, we can only estimate the difference between the skin and the bucket temperatures from the net energy flux $Q^*$ from the sea to the atmosphere (see Kraan, 1977).
If it was assumed that the thickness of the viscous skin layer is the same as the thickness of the temperature skin layer (in agreement with Saunders (1967) and Hasse (1971)), we have, taking $C_D = 1.25 \cdot 10^{-3}$:

$$T_w - T_a = \frac{v_w}{k_w} \sqrt{\frac{\rho_w}{\rho_a}} \frac{Q^*}{C_D} \frac{1}{u_{10}^{-1}} \simeq 0.014 Q^* (\bar{u}_{10})^{-1} \quad (11)$$

Here $\lambda$ is a constant in the order of 5 to 10, and $\bar{u}_{10}$ is the average wind velocity 10 m above sea level. $Q^*$ is the sum of the sensible-heat flux $Q_{Ha}$, the latent-heat flux $Q_{La}$, and the net flux of long-wave radiation $Q_{ir}$ (see Fig. 7), which can be found from the bulk transfer equations (Kitaygorodskii et al. 1973, and Hicks, 1975):

$$Q_{Ha} = \rho_a C_{pa} C_H \bar{u}_{10} T \simeq 1.9 \bar{u}_{10} (T_0 - T_a(10))$$

$$Q_{La} = \rho_a L_v C_L \bar{u}_{10} q \simeq 4.7 \bar{u}_{10} (q(T_a) - q(T_a(10)))$$

$$Q_{ir} \simeq \sigma T_a^4 (1 - (0.53 + 0.067 \sqrt{e})(1 + 0.13c) + 4 \frac{T_0 - T_a}{T_a})$$

Here the sensible and latent bulk transfer coefficient in air $C_H$ and $C_L$ respectively are taken $1.45 \cdot 10^{-3}$. These flux expressions are valid only for a near neutral atmosphere, and this generally is not the case with fair-weather conditions, under which most thermal mapping is carried out.

Figure 8. The diurnal variations of the sea surface temperature $T_o$ and the surface layer temperature $T_w$ at 20 cm depth. Mixing takes place in the surface layer.
If one is interested in the physics of the air-sea interface itself, the skin temperature is very useful, e.g. for calculating the evaporation. Fig. 8 shows the diurnal variations of the sea surface temperature $T_o$ and the temperature of the sea surface layer $T_w$ at a depth of 20 cm (the layer where intense mixing takes place).

2.4.3 Surface pollution of the sea

In reality, the surface of any open body of water is contaminated with floating natural or synthetic surface-active substances (surfactants). However, since it is very difficult to sample or even to detect these, it is almost impossible to study their effect on the data merely by calibration and measurements on the spot. Their most likely effect is a change in the refractive and reflective index. It would therefore be useful to study the emission coefficients of surface-active substances found on the sea surface. Pollution by aliphatic alcohols is known to reduce evaporation by 40%, raising the skin temperature by about 0.3 K.

It would be desirable to detect oil slicks at sea by measuring the radiation temperature, but their behaviour is complicated and varies with time, place, and the composition of the oil. Details are discussed in Section 3.4.2.

2.4.4 Variations of the sea surface temperature

The sea surface temperature varies both with time and with place. The temporal variations are due to seasonal effects, differences between day and night, tides, currents, estuarine or river outflows, weather effects (energy fluxes), atmospheric influences on the signal, and biological phenomena, such as natural surface-active substances. The spatial variations arise because of differences in the depth, distribution of mass, pollution, cloud cover, tidal currents, outflow and upwelling.

The magnitude of the resulting variations in the sea surface temperature likewise depends on these factors of time and place. In the case of airborne measurements over fairly long periods both types of variation are present, which may result in non-synoptic composites.

On small spatial and time scales (100 m, 120 s), and in fair weather, the radiation temperature can vary as much as $\pm 3$ K, owing to the small heat capacity of the thin skin layer. Scattered cumulus clouds can induce a 'noise' of $\pm 0.5$ K.

In the case of outflows from estuaries and harbours, it is remarkable how a clear distinction between 'old' and 'new' freshwater persists for many tidal cycles and over
long distances (80 km). It seems that small differences between the density of freshwater and seawater strongly inhibit the mixing between them. However, sharp boundaries between them do not necessarily coincide with isotherms.

2.5 Radiation temperature of the ground

2.5.1 Factors that exert an effect

In theory, the surface temperature of the ground or of the plants and objects on it can be determined from the amount of thermal radiation received by the sensor. In most applications we are interested not only in the surface temperature but also in the characteristics of the surface that exert an effect on this temperature. For an accurate determination of the radiation temperature of the surface, we must determine the proportion of energy that is lost by scattering and absorption in the atmosphere, together with the energy contribution of the atmosphere in the same wavelength region (see Section 2.3.7). The radiation emitted by the soil surface (which here will include vegetation, buildings, and other objects on it) is transmitted to various extents, depending on the atmospheric conditions. The characteristics of the surface itself also play a role. For the sake of simplicity it is generally assumed that the surface behaves as a black body radiator. An emission coefficient of less than 1 is sometimes used, on the assumption that the surface behaves as a grey-body radiator, i.e. that $\varepsilon$ is independent of $\lambda$. Such assumptions can lead to fairly large deviations, because the energy is always detected in a narrow frequency band ($\lambda_1, \lambda_2$) in thermal infrared measurements. The emission coefficients generally used mostly relate to the total energy radiation (see Fig. 9), and differences of e.g. 1% are ignored. However, even such a difference in the emission coefficient leads to a 0.6–0.7 K error in the determination of the radiation temperature under a clear sky (0.1 K differences are usually regarded as still measur-

Figure 9. A typical diagram for the reflected solar radiation and the emitted thermal radiation. Note the overlap in the 3-5 μm band.
It is evident that, if the emission coefficient in the frequency band \((\lambda_1, \lambda_2)\) used differs from the (energetically) averaged emission coefficient, the error in the determination of the radiation temperature becomes much greater still. If there is also an angular dependence of the emitted energy (see Section 2.4.1), the deviation will be even greater.

2.5.2 Structure of the vegetation cover

The vegetation cover often consists of two or more layers of different plants one over the other; only arable and market-garden crops consist of just one type of plant.

The density of foliage varies along the height of the vegetation cover: as we go down from the top, the vegetation becomes thicker and thicker. Sometimes it then decreases, possibly only to increase again, forming another layer of vegetation, e.g. of moss under a layer of grass in a meadow. The transition to the solid ground is often difficult to specify, as it consists of plant debris in various stages of decomposition, with mineral particles incorporated in-between. Gradually there is more and more soil, and finally we come to the roots. The vertical distance between something that is definitely above the ground and something that is definitely in the ground may range from a few millimetres to a few centimetres.

The primary cause of all temperature differences is the incident solar radiation, known also as insolation. The radiation received at the ground comprises light in the visible region (see Fig. 9). This short-wave and near-infrared radiation impinges on the plants. The top leaves reflect some of it back, absorb some of it, and let the rest through. The lower leaves do the same, but they receive not only direct sunlight and the diffuse sunlight scattered in the air, but also radiation transmitted by the higher leaves and reflected radiation coming from both higher and lower leaves. If the vegetation is dense, it can happen that the soil itself receives no direct sunlight or diffuse sky radiation. The radiation absorbed by the leaves is used by the plants for heating and evaporating water in their various parts and for a minor part for the chemical processes of assimilation.

The extent of heating and evaporation depends on several processes. All leaves emit thermal radiation at a wavelength between 3 µm and about 0.1 mm. The leaves also receive thermal radiation from all sides. In the case of the top leaves this radiation comes from the atmosphere above, from the other leaves and plant components below, and possibly even from the bare ground. As we go lower down, the leaves receive less and less radiation from the atmosphere but more and more from the leaves above them. They still receive radiation from below, coming from lower plants as described in the case of the top leaves.
Since in the latitudes of the Netherlands and in higher latitudes the atmospheric radiation temperature is generally lower than that of plants (even at night), the plants lose energy by this thermal radiation, and the higher their position in the vegetation cover the more energy they will lose. The amount of energy used for assimilation is almost negligibly small in comparison with the amount of energy used for other purposes. So it may be stated that plants use solar radiation mainly for evaporation and heating.

Another process consists of the heat exchange with the surrounding air. Leaves fully exposed to the sun, which receive the largest amount of energy, become warmer than the air around them and so lose heat to the atmosphere. The more intense heating also means that these leaves will evaporate more water, while lower leaves are heated to a smaller extent. If there is wind, air penetrates the plant canopy, and the lower leaves are heated by the warmer air. As the temperature rises, the leaves will radiate more heat.

Since an infrared radiometer directed at a plant from above receives that part of the radiation emitted from the leaves which is not intercepted by other leaves. The top layer per square meter leaf area contributes more to the input signal than do the lower parts of the plant. Of course, they also contribute more to the heating of the surrounding air, but the lower layers of the vegetation cover contribute more to heat exchange than the top layers, particularly when there is any wind. The situation is different with evaporation: if there is wind, heat penetrates downwards and the upward flux of water vapour is increased by the wind. The air temperature in the lower layers of the vegetation cover therefore clearly plays a very important role.

The ground also receives and releases heat by other mechanisms, mainly by conduction. The amounts involved depend on the nature of the soil, but are generally much smaller than those in the processes described above. Yet this heat exchange does have an effect on the plant, influencing the leaf temperature and thus the radiation temperature. In fact, we try to deduce something about the soil from these rather small differences in the radiation temperature, and surprisingly enough, this can be done.

Two other indirect characteristics that are of importance should be mentioned at this point. One is the ratio between the energy used for heating the air and the energy used for evaporation (Jacobs et al., 1977). This is known as the Bowen ratio and it depends on the wind to an extent that increases with increasing dryness of the air. A stronger wind means a lower leaf temperature, less heating of the air, less thermal radiation, and more evaporation.

The second effect is due to a physiological reaction of the plant. If a plant does not get enough moisture from the soil, it closes its stomata so that evaporation is reduced. This raises the leaf temperature, which in turn leads to a more intense heating of the surrounding air and to more thermal radiation. This effect can intensify the above heat
exchange with the ground, but it can also mask it or even reverse it. Solar irradiation begins about half an hour before sunrise and increases from zero to a maximum, which is reached around noon (solar time) if the sky is clear. It then gradually falls off to zero some time after sunset. This produces diurnal cycles in all thermal processes (see Fig. 10). The processes of heat exchange and heat utilization described above are not in phase with the insolation cycle, the result being that a warmer surface sometimes cools sooner than a colder one. In practice, e.g. a difference in the grass cover can bring about phase shifts of up to several hours because of a different ground heat flow. The ‘warm’ surface can then get colder than the initially ‘colder’ surface in the course of the afternoon.

![Figure 10. A typical diurnal cycle of the soil radiation.](image)

During summer almost every influence of the soil is masked by the vegetation cover of the fields. Corn has completely different characteristics e.g. from sugar beet. In pastures we have fields with freshly mown scattered hay, fields with grass grazed bare, fields with meadow grass, and fields with much longer grass for winter fodder. Furthermore, owing to hydrological differences in the same field, there will be different types of vegetation. Identification of the grasses can provide information about all the wet parts of grassland. Besides, differences in tilling and fertilization can introduce various finer differences.

The ‘ground heat model’ constructed in the Appendix takes into account the various effects discussed above. A comparison with the models of Bonn (1977) and Kahle (1977) who assume a simple surface, shows the advanced character of our model. Kahle’s model was developed for bare ground, and Bonn’s model for bare ground with grass cover. Our own model partly explains the large scatter of Bonn’s measurements. The regression line obtained by plotting the ground surface temperature as a function
of the radiation temperature, found by Bonn, also includes the phase shifts, which is not explained by him.

2.6 Characteristics of radiation thermometers and scanners

Photographic emulsion is insensitive to thermal radiation. The thermal radiation is monitored with a solid-state detector that measures the incident photon current. At room temperature the internally generated current ('noise') limits the detection of very low energy levels, so that it is difficult to detect small temperature contrasts corresponding to very small energy differences. To improve the signal-to-noise ratio, the detector is cooled to the boiling point of liquid nitrogen (62.5 K) or even that of liquid helium (1 K).

The detector is used at room temperature in infrared radiation thermometers ('radiometers') to measure the absolute radiation level. The sensitivity now can be improved by integrating the incoming radiation of a given area by means of measuring over a longer period. If these radiometers are carefully calibrated, they have an absolute accuracy of ±0.5 K and a relative accuracy of ±0.05-0.02 K, depending on the type of instrument used. All IR radiation thermometers give either point readings or temperature profiles.

Cooled detectors are used in infrared line scanners (IRLS) employed in airborne scanning over large areas. These instruments produce what is known as a thermal image (or imagery) of the area covered. With accurate calibration, the thermal resolution is about 0.2-0.1 K. The simultaneous use of infrared radiation thermometers and infrared line scanners gives calibrated thermal maps. We shall describe some of the instruments here, but for further information the reader is referred to Hudson’s book (1969).

It may be mentioned that the thermal band is usually included in commercially available multispectral scanners (for a detailed description, see Reeves, 1975). It is in fact useful to combine infrared data with information obtained in the visible window when studying e.g. the vegetation, hydrology, oil spills, geomorphological structures, and coastal phenomena.

2.6.1 Infrared radiation thermometers

These instruments use a mechanical chopper to compare the amount of incident radiation with the amount of radiation from a calibrated internal source at a convenient temperature, generally 48 °C.

The resulting energy difference is a means of measuring the incident radiation
temperature which has to be calibrated. The absolute object temperature is then calculated (see Section 2.3) taking into account the atmospheric and emission effects. The characteristics of some typical infrared radiation thermometers are given in Table 2.

2.6.2 Infrared line scanners

These instruments use a rotating mirror with an instantaneous field of view with an angle \( \beta \) in both directions (see Fig. 11). Film or magnetic tape is used to record the signal.

The mirror scans the object through an angle \( \theta \) in a plane normal to the track of the airborne platform. The speed of rotation of the mirror is adjusted to the speed at which the platform is travelling, in such a way that successive scan lines touch one another or partially overlap (see Fig. 12). This technique is called linear scanning. When a linear time base is used in this system, the image is distorted at the edges because of the projection from the flat plane onto a semi-cylindrical surface (see Fig. 13). To obtain a geometrically correct image, and also because of various movements of the platform.
Figure 12. Geometry of an individual swath (A) and the successive coverage of the area (B).

Figure 13. The cylindric projection of the flat terrain on the image plane. Note the large distortion towards the edge of the image plane.
Figure 14. General distortion of the image due to the movements of the platform. I flight configuration, II image plane: a image plane, b actual coverage.

Table 3 Characteristics of the HBR-Singer Reconofax-VI Infrared Line Scanner.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band-width of the video signal</td>
<td>0-200 kHz</td>
</tr>
<tr>
<td>Wave-band pass</td>
<td>8-14 μm</td>
</tr>
<tr>
<td>Scanning speed</td>
<td>400 lines s⁻¹</td>
</tr>
<tr>
<td>Scanning angle</td>
<td>140 deg</td>
</tr>
<tr>
<td>Instantaneous field of view</td>
<td>3 mrad</td>
</tr>
<tr>
<td>Thermal resolution</td>
<td>0.15 K</td>
</tr>
</tbody>
</table>

(i.e. pitch, roll, yaw, and drift) and changes in its velocity and altitude, a point-to-point reconstruction must be carried out. Fig. 14 shows the effects of aircraft yaw and drift, and Table 3 sums up the characteristics of the Singer Reconofax-VI instrument used in this research.

2.6.3 Infrared conical scanner

In the case of infrared line scanners, as described above, the geometric resolution and the angle of incidence vary along the scan lines (see Fig. 13). To eliminate these variations, the Physics Laboratory TNO designed and constructed a conical scanning system in which both these parameters are kept constant (see Fig. 15). The fact that the sensor performs a full revolution enables two successive observations of the same object to be obtained. The instrument is designed to detect very small temperature differences between fairly large objects: the thermal resolution has been improved at the expense of geometric resolution. Because of the narrow temperature
range involved (1–3 K), this is in fact useful for studying the temperature distribution over large water surfaces and over land for the detection of e.g. large sand veins or fossil river beds. Since the fixed angle of incidence does not introduce other variables, the emission coefficient and the nature of the reflection from the water surface can be studied in detail, especially when the two atmospheric windows (3–5 and 8–14 μm) are used simultaneously (see Section 2.2.1).

2.7 Surface truth measurements

As we have seen, theoretical relationships can be used to determine the influence of the atmosphere and the reflections from the surface. In practice, however, these relationships are difficult to verify owing to the stochastic and changing character of the various factors involved. The corrections we have derived are valid if certain assumptions have been fulfilled, e.g. if the temperature and humidity profiles are in fact uniform.

To understand the fluctuations better, airborne measurements are combined with simultaneous in-situ measurements on the surface, which of course are free from any atmospheric effects. The reflection is determined as the difference between the observed directly incident radiation and the upward radiation. In principle, the simultaneous measurements need only be done at the time of the flight, but this does not enable one to study the heat-exchange processes. A remote sensing observation from a moving platform only gives a ‘snapshot’ in time, whereas most heat exchanges occur over a
relatively long period (see Section 1.2). For the sake of a better study and understanding of the physical processes, one must therefore conduct prolonged and detailed measurements over a limited area. This gives both the necessary data for the time of the flight, and information about the changes that occur prior to the flight. The way a certain situation comes about in the heat balance is often important in the interpretation of the images obtained. We shall therefore discuss in more detail the various surface truth measurements generally carried out in this work. In Section 3 we shall also discuss the additional special measurements required at the various measuring sites.

2.7.1 Surface truth measurements on land

2.7.1.1 THE FIELD STATION

The field station for surface truth measurements consisted of a weather station, a series of extra thermometers, recording radiometers, and sometimes a limnimeter to measure the ground-water level (see Fig. 16). The data were digitally recorded on magnetic tape. The weather station was equipped with sensors (Monteith, 1972) to measure the wind velocity and wind direction at a height of 2 m, the air pressure, the temperature and relative humidity of the air at a height of 1.5 m, the incident solar radiation and the albedo (Kipp albedometer CM 7), the precipitation, and the turbulence (sonic anemometer developed by Kaiyo Denki, PAT-112-1 (Mitsuta, 1963; Van den Abeele et al., 1976)). During the various experiments the station was developed to incorporate all relevant parameters. Both, albedo and turbulence, for example, were included in the last experiment (see Sect. 3.1.2).

OTHER SENSORS

Additional temperature-sensing devices were used to record the temperature between a seepage site and its immediate surroundings at depths of 10, 20, 40, and 100 cm. Two or three sets of nickel resistance thermometers connected in series were used to measure the temperature gradient over the surface (at +3, 0, -3, and -6 cm).

Three types of radiometric measurements were done at a number of places:
1 First the net radiation was measured with a Funk net radiometer (Funk, 1962), with a bandwidth of 0.3–50 μm. The instrument was kept free from dew by aspiration (Van den Abeele, 1976).
2 The radiation temperature was also measured with an absolute radiation thermometer developed by the Physics Laboratory TNO based on Stoutjesdijk’s design (1966). The incoming radiation is measured with two identical sensors built in the same copper housing, but with different bandwidth filters. One is screened by a
Figure 16. Several sensors of the field station used for surface truth measurements. 1 weather station, 2 temperature profile sensors (up to 1 m depth), 3 solar radiation sensor, 4 absolute radiometer, 5 resistance thermometers for temperature gradient measurement (up to 6 cm depth), 6 sonic anemometer.
polythene film with a 0 to \( \infty \) bandwidth, the other by glass with a 0.3–2.5 \( \mu \)m bandwidth. The difference of both signals, corrected for the housing temperature, is recorded. The housing temperature is measured with a nickel resistance thermometer, which is described above. The resulting signal correlates with a Barnes PRT-5 radiometer reading to 0.99, with a standard deviation of 0.3–0.6 K. This enables the radiation from 2.5 to \( \infty \) to be measured and corrected for the ambient radiation within the housing, which affects the radiation sought. The albedo within the 0.3–2.5 \( \mu \)m is canceled out by subtracting the two signals measured (see Sect. 2.5.2, Fig. 9).

3 We also measured radiometrically the temperature differences where temperature contrasts were expected. This was done with two or more groups of radiometers, connected in series, these being of a simpler type than the ones described in 2 above. Each comprised only one chamber, which was covered with a polythene film, a material that transmits more than 90\% of the thermal infrared radiation. Starting with the assumption that the reflectance, or albedo, is the same at both measuring points, we could interpret the differences observed between the sets as temperature differences. This does not always apply in practice, even in the case of a thin grass cover, and much care is needed in the interpretation of the observations.

2.7.1.2 RESULTS

The variables mentioned above were all recorded continuously during the measuring period, which generally lasted a few months. These records were used in the direct interpretation of the thermal images obtained, but they were also intended to provide an insight into the micro-meteorology and the heat balance of the test area. The following analyses were carried out in addition to the standard analyses discussed in Section 3:

a The variations of the average temperature as a function of depth were determined (see Fig. 17). The graphs indicated the expected temperature contrast at sites where direct seepage occurs.

b The damping depth is defined as the depth at which the temperature variation has fallen off by a factor 1/e. The damping depth is approximated as follows. In a half infinite homogeneous medium the amplitude of the diurnal temperature cycle must decrease exponentially with increasing depth (Van Wijk, 1966; Monteith, 1975). The damping depth then can be estimated with sufficient accuracy by taking the temperature variation as a function of depth. Fig. 18 shows an example, for measurements carried out near the river Waal in 1976–77 (see Section 3.1.2). It can be seen that two layers must be distinguished, i.e. a top layer extending down to a few centimeters, and a deeper layer. This situation occurred in all the measurements. Where the soil has a two-layered structure the approximation is applied twice in succession. The damping depth was thus determined for the top layer and the deeper
layer in all the measuring periods, always as the mean value for a few days during which the conditions were comparable.

Figure 17. Average daily temperature profiles for the near-surface soil, on different dates (— x — field II. — field I, experiment near the river Waal 1976/77).

c When the remote radiometric recordings were compared with the in-situ temperature measurements at levels of 0 and +3 cm above the surface, it was found that as the vegetation grew its influence increased, eventually becoming dominant. In the case of a thin grass cover in winter the mean of the temperatures measured at 0 and +3 cm was in good agreement with the radiation temperature recorded, the correlation between them being about 0.99. Here it was even possible to find the height of the reference plane for the radiometric measurements with a reasonable accuracy.

d An accurate analysis of the observed diurnal temperature cycle often revealed a phase shift between the areas between which a temperature contrast had been expected.
Figure 18. The damping depth $D$ for the seepage field and the reference field, based on measurements such as shown in Figure 17 (Experiment near the river Waal 1976/77).
Van Wijk (1966) mentioned these small phase shifts in the case of areas with dissimilar thermal characteristics (see Chapters 4 and 6 of his book). Small differences in the top layer greatly intensify this effect (see Section 2.5.2). We regularly observed differences of up to 20 min during the experiments. This has important consequences for the time at which the thermal image shows the maximum contrast. In the absence of a phase difference of this kind in the diurnal temperature cycle the maximum contrast is expected to occur around noon, but a phase difference of 15–20 min shifts this maximum by several hours. In the case of the measurements near the river Waal (see Section 3.1.2), this shift is retrogressive, towards the morning.

2.7.2 Surface truth measurements at sea

In principle there is no difference between land-based and sea-based measurements as regards the parameters to be measured at zero height, but at sea it is generally impossible to set up a long-term measuring station at the required location. Measuring buoys have so far proved too limited, and as a rule ships and observation towers have to be used. Yet the same requirement applies to remote sensing over the sea, namely that there should be at least one surface station in the area scanned. In practice, this mostly means that the area is so chosen that it contains a marine observation post (either a lightship or an observation tower), though in practice the latter is never equipped with the complete set of instruments found in a land-based station, owing to the lack of space, the problems of sustenance, and above all the harsh environment. A minimum observation programme should however include the following measurements:

- concerning the sky: cloudiness and cloud-base temperature (to be measured with the same radiation thermometer)
- at flight altitude: air temperature
- at 10 m: wind velocity, air temperature and humidity
- at sea level: height of the waves, presence of surface-active substances
- at a depth of 1 m: bulk temperature of the water

It is important to extend this programme, and in particular to include measurements of the net short-wave radiation and the atmospheric stability. It is also advisable to carry out, at the beginning and end of the airborne remote sensing missions over the sea, a number of measurements at a lower altitude over a sea area showing the greatest possible homogeneity of the surface temperature, or else to perform simultaneous measurements at sea level with the aid of a second, identical, radiation thermometer. It is not possible to compare directly the radiation temperature (skin temperature) and the ground truth (bulk temperature), but the two can be compared by determining the temperature difference over the surface layer with the aid of Eq. (11) (see Section 2.4.2). Fig. 19 shows the vertical sea temperature profile with the calculated correction for the
radiation temperature. Except after prolonged periods of absolutely calm weather, the water below the surface layer can be assumed to be fully mixed. On the other hand, we must bear in mind the limited applicability of the equation, the variations in the emission coefficient due to patches of surface-active substances, and variations in the surface temperatures with time and place (see Section 2.4.4).

Figure 19. Temperature profile of the sea with the corrected radiation temperature.

2.8 Image enhancement

The digital processing of images comprises two-dimensional transformation, image reconstitution, the use of filters, segmentation, and analysis of image parts. Advanced techniques have been described such as by Shanda (1976) and by Simon and Rosenfeld (1977).
What we needed in this work was mainly image enhancement, aimed at improving the contrast by reducing the noise level in the image and at determining the absolute temperatures in the image. This processing gives a thematic map of the absolute temperature distribution over the area scanned.

Figure 20. a) Original thermal map taken with a Reconfax line scanner at about 12.30 p.m., January 31st, 1977. Black is warm and white is cold.
b) The same scene digitally processed, showing dynamic resolution improved at the expense of the geometric resolution. Black is warm and white is cold.

The noise level in the image can be reduced in various ways, and we chose spatial averaging, in which some of the geometric resolution is sacrificed in return for a higher thermal resolution. The temperature can be calibrated with the regression line obtained by plotting IR radiometric data against the IR line scanner data, and by using surface truth measurements for known reference points in the image (e.g. water surfaces).
We mostly used the film-recording technique, but some of the records were processed digitally. The image processing was carried out as follows: The video signal from the ReX scanner (see Section 2.6.2, Table 3) had a band-width of 0-200 kHz. The signal was passed into a recorder, which had a band-width of only 0-40 kHz; this signal was sampled at a frequency of 80 kHz. With a period of 2 ms per scan line, the number of image points amounted to 160.

Picture elements (pixels) at the edges of the observation angles were ignored, because they give strongly distorted information. The use of 120 image points then corresponds to an angle of observation of 90°. The apparent aperture is then $90°/120 = 0.75°$ or about 13 mrad, instead of 3 mrad, which is the instantaneous field of view of the infrared line scanner itself. When an airborne instrument flown at an altitude of 300 m is pointed vertically downward, the scale is such that each pixel corresponds on the ground to an area with a diameter of 4 m. To obtain rectangular pixels all the image lines produced while the aircraft moves through 4 m must be combined, which at a flight speed of 60 m/s amounts to 33 lines. The noise level in the image was reduced by a factor of $\sqrt{(200/40) \times 33} = 13$ by using the reduced video band-width and combining a number of scan lines. The thermal resolution of the scanner was about 0.3 K, which was improved to 0.025 K by the technique described. This improvement can be clearly seen in Fig. 20: the digital image processing shows up thermal differences previously hidden by the noise in the film.
3 Applications

All locations to be mentioned in the following are shown on the map in Fig. 21.

Figure 21. The locations of the test sites mentioned in Chapter 3.
3.1 Hydrology

Thermal mapping with an infrared line scanner was done over three sites in the Netherlands, all with similar hydrological problems. In all three cases a high water level in a river or a canal forced water to seep through the bed of the waterway into the adjacent land. In one case salt water has caused damage in lower-laying adjacent agricultural land. In the second case the stability of a river dike was threatened by a large runoff after the winter period. The third case concerned seepage of water from a canal section between two locks which meant that the water balance could not be estimated by sole reference to the gating. Clearly this has important consequences for shipping.

3.1.1 Seepage of salt water into agricultural areas

The Ghent-Terneuse canal in Zeeland allows sea-going vessels to sail up to Ghent in Belgium. The water in this canal has a high salt content, partly from water entrained with the sea-going vessels as they move through the locks and partly from industrial discharges along the canal banks. The water level in the canal is constantly 2 m above the ground surface in the polders. The canal cuts through an alluvial area, where the last polderization occurred in the 1960's. The inhomogeneous soil consists of young marine clay alternating with packets of peat, layers of sand, and sanded-up creeks. Unlike clay and peat the last two are permeable to water. The shallow ground water is salty, containing 300-10,000 mg of Cl$^-$ per litre. Where the covering clay layers are cut through or are very thin the salt water surfaces, producing a stress situation for the vegetation present.

By thermal mapping we located the seepage sites along the canal and measured the size of each site. We also used false-colour photography, which clearly shows up differences in the crop growth (see Fig. 22, taken over the Autriche polder). To understand the mechanism of seepage and the differences in the radiation temperature, we measured the ground temperatures and the meteorological variables at a field station (see Section 2.7.1), located at an established seepage site in the Autriche polder in the period between 6th June and 8th August 1975.

When plants are distressed they develop differently and as a consequence reflect visible light differently (Bunnik, 1978), making the colours in the false-colour photograph lighter and possibly even white (see Fig. 22, and Reeves (1975) for a detailed description and interpretation of the false-colour technique). These plants also have a completely different evaporation pattern (see Section 2.5.2), as a result of which the thermal balance of the leaf is different from that found in healthy plants. This gives rise to an anomalous radiation temperature, which shows up in the thermal image. Fig. 23 shows
Figure 22. False colour photograph of the test area in the Autriche Polder taken at 11.15 a.m., June 9th, 1975. The seepage vein and the areas affected by seepage are light coloured, the other areas being dark red and dark green. The old creek system in the centre is discernable clearly.
the temperature contrast over a 24-hours period, between the seepage site and its surroundings with healthy vegetation. The phase shifts in the radiation temperature curves for the seepage site and its normal surroundings could only be explained after further experiments (see Section 3.1.2).

![Graph showing temperature contrast cycle](image)

Figure 23. Mean temperature-contrast cycle between the seepage site and the surroundings, over a 24-hours period, in the Autriche polder.

Since there was very little rain during the test period, the thermal conductivity characteristics of the ground did not show any great changes. The damping depth $D$ measured for the seepage site and for its surroundings are shown in Table 4. Only the changes occurring during the summer season were involved.

| Time, 1975 | Diurnal damping depth D, cm |
| --- | --- | --- |
| 10-13/6 | field II | field I |
| deep | 11.5 | 14 |
| top | 6 | 5.2 |
| 30/6-7/7 | deep | 12.5 | 13.3 |
| | top | 5.5 | 6.8 |
| 30/7-3/8 | deep | 22.5 | 12 |
| | top | 5 | 4.5 |

Unfortunately, in the autumn of 1974 and the winter of 1974/75 there had been some prolonged rainy periods, and the resulting excess water created a stress situation for the vegetation cover in some places, which masked the seepage of salt water. Moreover, as the seepage site in the Autriche polder was difficult to till, it was sown later than the rest of the area, so that the vegetation over it was retarded for this reason as well. To locate the seepage, therefore, we took soil samples in all suspect areas lying in a 200 m wide
strip along the canal and analysed them for their chloride content (see Fig. 24 for the Autriche polder). Only a significantly higher chloride content indicated the salt-water seepage.

To clarify the structure of the strata in the Autriche polder, a number of geoelectric measurements were carried out (see Fig. 25).

There was a clear agreement between the isohaline lines for a depth of 0.60 m below the surface (Fig. 24) and the electrical resistivity pattern at a depth of about 1 m, (~ NAP, Normal Amsterdam Water Level) (see Fig. 25).

![Figure 24. Chloride concentration distribution in soil layers at 0.2 and 0.6 m depth in the Autriche Polder. Concentration is given in mg/l.](image)

It could be assumed that the seepage centre was fed from the canal along a seepage path, but it was not known whether the seepage occurred through the side walls or through the bottom. The seepage stream was calculated using a greatly simplified simulation model to estimate its order of magnitude and to determine the source point, but no useful results were obtained, because of the lack of accurate data on the stratification and the properties of the strata.

The mean diurnal temperature differences between the seepage centre and the surroundings are moderate to large, ranging up to 5 K (see Fig. 23). As the vegetation emerged and grew, it screened the ground and determined the thermal behaviour of the
Figure 25. Geoelectric measurements indicating the resistance of the soil layers at several depths. The resistance is a measure for the chloride content and hence for the seepage.

surface, but the vegetation over the seepage site, which was sown later, was also retarded in its development. As a result, the seepage site both warmed up sooner and cooled more quickly at the beginning of the test period (May-June), but the situation
changed at the end of July and the beginning of August: the temperature then varied more sluggishly over the seepage site than it did over its surroundings. This can also be clearly seen in the thermal maps (see Fig. 26). On the occasions of the thermal mapping a fair to good contrast was always observed (see Fig. 23). The measured temperature difference of about 2 K was confirmed by the thermal map shown in Fig. 27, which has been processed digitally. It is interesting to note that the creeks clearly visible in the thermal map taken at 5:15 p.m. are comparable to those in the false-colour photograph (Fig. 22) taken at around 11:00 a.m. (Central European Time in both cases).

We found that, used in conjunction with false-colour photography, thermal mapping can be used to locate seepages and to estimate their extent. The resulting information was successfully used later for directing expensive field experiments.

3.1.2 Seepage of water due to high river discharge

When the river level is high at the end of the winter, dike-threatening seepage occurs along the river Waal, a branche of the river Rhine (see Fig. 21), between Tiel and Gorinchem. The area traversed by the river Waal is characterized by sedimentary clay alternating with sand beds and packets of peat. As the river changed its course in the past, meanders were cut off and sanded up, giving rise to a very heterogeneous geological structure. The dikes were partly built on these sanded-up meanders or on the sand veins.

At a high water level these sand veins and layers let the water through, and where the clay cover is very thin or is dissected the ground water appears on the surface. The location of the seepage springs shows a somewhat different pattern from year to year, depending on the river discharge. New springs may be added, and these changes have to be recorded every year.

In winter the temperature of the upwelling ground water is usually different from that of its surroundings, and therefore an attempt was made to use thermal mapping to locate the seepage sites and the sand veins.

Such exercises have to be carried out over limited time periods, since the high water level in the river lasts for 2–4 days only and the resistance to the ground water flow is too high to reach a stable situation. Besides, the weather conditions in this period must be suitable for thermal mapping, i.e. there must not be any rain or fog. During the period in question (December 1976 to March 1977) there were three peaks in the river level (above the Normal Amsterdam Water Level): 4.80 m at the end of January, 6.00 m in the middle of February, and 5.50 m at the end of February, the polder water level being 0.10–0.30 m (above the NAP). Snow also fell in this period, and there was a frost lasting for weeks.
Figure 26. Thermal maps of the test area in the Autriche Polder.

a) taken at 10.30 a.m., June 10th, 1975.
b) taken at 5.15 p.m., August 6th, 1975.
c) taken at 5.30 a.m., June 10th, 1975.
d) taken at 11.00 a.m., August 1st, 1975.
Figure 27. Digitized thermal map of the Autriche Polder test site, August 6, 1975, 1700 h C.E.T.

Owing to the phase shift in the temperature cycle between the ground water and the surface, the snow cover and the frost were very useful in the detection of the seepage
springs, for they give the greatest possible temperature contrast (5–10 K). Since the seepage stream was continuous, it did not cool quickly, so that the uncovered springs could be detected very easily from the air (See Fig. 28).

Figure 28. Thermal maps of the test site near the river Waal at two different moments.
1. Land has a higher temperature than the water due to the insolation during the day.
2. Land has a lower temperature than the water due to the nocturnal radiation of the heat.

The list of springs detected between Tiel and Gorinchem with the aid of the thermal images was compared with the results of simultaneous field exploration. The former technique showed up 40 sites and the latter 26, only 8 coinciding in the two cases. However, the following points should be borne in mind:
- we were dealing here with a constantly changing phenomenon (a higher water level gave rise to more seepage springs)
- the view from the aircraft was sometimes obscured by trees, buildings, etc.
it was possible that inadvertently more attention was paid to covered springs in the field exploration.

This comparison illustrates the supplementary character of thermal mapping.

The sand veins lying near the surface can also undermine the stability of the dikes. A high soil moisture tension can suddenly give rise to a spring. These sand veins can be detected only in the absence of a snow cover, because the temperature differences are very small. To detect such sand veins by thermal mapping, the field station (see Section 2.7.1) was set up near a previously identified sand vein in an area known as Het Hoogland near the village Herwijnen. Here temperature, the ground-water level, and the usual meteorological data were recorded in the period between 6th December 1976 and 15th March 1977.

Table 5 shows the damping depths measured for the spot with sand below the surface (field II) and for its surroundings (field I). This period was chosen because the evaporation is slight, the trees are bare, the grass in the pastures is almost dead, the farm animals are in their stables, and the arable land is free from vegetation. All these factors favour the detection of both seepage springs and sand veins.

Table 5  Damping depth measured at the river Waal in the winter of 1976/1977. Field II is the clay-covered sand vein and field I is its surroundings

<table>
<thead>
<tr>
<th>Period</th>
<th>Field I</th>
<th>Field II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>top layer</td>
<td>deep layer</td>
</tr>
<tr>
<td>1976</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-13/XII</td>
<td>3.5</td>
<td>14.5</td>
</tr>
<tr>
<td>1977</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-19/I</td>
<td>3.0-3.7</td>
<td>19</td>
</tr>
<tr>
<td>29/I-2/II</td>
<td>1.5</td>
<td>-</td>
</tr>
<tr>
<td>4-6/II</td>
<td>3.5</td>
<td>-</td>
</tr>
<tr>
<td>8-14/II</td>
<td>3</td>
<td>13.5</td>
</tr>
<tr>
<td>13-17/II</td>
<td>2.8</td>
<td>13.5</td>
</tr>
<tr>
<td>22 + 23/II</td>
<td>3.2</td>
<td>13.5</td>
</tr>
<tr>
<td>7-14/III</td>
<td>2.5</td>
<td>12.5</td>
</tr>
</tbody>
</table>

The radiation-temperature cycle for the sand vein and for the reference field was also calculated with the aid of a 'soil heat model' (see Appendix), and a first order approximation gave good agreement with the measurements (see Fig. 29; cf. Fig. 10 for the thermal flux). The calculations indicated that the vegetation cover has a dominant effect on the heat balance and thus on the radiation temperature. This fully explains the phase differences detected in the measurements.
In contrast to the case discussed in Section 3.1.1 (see Fig. 23), the maximum temperature contrast occurred not at 12:00-2:00 p.m. (found for a salt water seepage in the summer) but earlier, i.e. at 11:00-12:00 a.m. (see Fig. 30). The position of the maximum is determined entirely by the phase differences. Table 6 gives the times at which the maximum temperature was recorded for the two places for which the damping depths are given in Table 5.

Table 6  Time of day when different methods indicated the maximum daily temperature at the river Waal in two periods. T(-3) = temperature at −3 cm; NETRAD = Funk’s net radiation meter; ABSRAD = radiometer built by the Physics Laboratory of the TNO (see section 2.7.1). Field II is the clay-covered sand vein and field I is its surroundings

<table>
<thead>
<tr>
<th>Sensor</th>
<th>1976</th>
<th>1977</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dec. 9–12</td>
<td>March 8–10</td>
</tr>
<tr>
<td>Solarimeter</td>
<td>12 hr 16</td>
<td>12 hr 48</td>
</tr>
<tr>
<td>Albedometer</td>
<td>12 hr 20</td>
<td>12 hr 52</td>
</tr>
<tr>
<td>Air temp.</td>
<td>13 hr 44</td>
<td>16 hr 44</td>
</tr>
<tr>
<td>Field I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T(+3)</td>
<td>13 hr 24</td>
<td>14 hr 56</td>
</tr>
<tr>
<td>ABSRAD</td>
<td>13 hr 32</td>
<td>15 hr 04</td>
</tr>
<tr>
<td>NETRAD</td>
<td>12 hr 40</td>
<td>13 hr 00</td>
</tr>
<tr>
<td>T(0)</td>
<td>13 hr 56</td>
<td>15 hr 16</td>
</tr>
<tr>
<td>T(−3)</td>
<td>15 hr 28</td>
<td>17 hr 00</td>
</tr>
<tr>
<td>T(−6)</td>
<td>16 hr 56</td>
<td>18 hr 08</td>
</tr>
<tr>
<td>T(−10)</td>
<td>18 hr 24</td>
<td>18 hr 56</td>
</tr>
<tr>
<td>T(−20)</td>
<td>0 hr 52</td>
<td>21 hr 04</td>
</tr>
<tr>
<td>Field II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T(+3)</td>
<td>13 hr 20</td>
<td>14 hr 56</td>
</tr>
<tr>
<td>ABSRAD</td>
<td>13 hr 20</td>
<td>14 hr 52</td>
</tr>
<tr>
<td>NETRAD</td>
<td>12 hr 32</td>
<td>12 hr 56</td>
</tr>
<tr>
<td>T(0)</td>
<td>13 hr 36</td>
<td>15 hr 00</td>
</tr>
<tr>
<td>T(−3)</td>
<td>15 hr 12</td>
<td>16 hr 32</td>
</tr>
<tr>
<td>T(−6)</td>
<td>16 hr 36</td>
<td>17 hr 44</td>
</tr>
<tr>
<td>T(−10)</td>
<td>18 hr 08</td>
<td>18 hr 28</td>
</tr>
<tr>
<td>T(−20)</td>
<td>0 hr 40</td>
<td>21 hr 12</td>
</tr>
</tbody>
</table>

3.1.3 Seepage of water into sedimentary basins

Comparable studies have been carried out by de Loor (1970) and de Hoop (1977) in the Alblasserwaard polder near Gorinchem. These similarly included both thermal mapping and observations with the field station described in Section 2.7.1. After preliminary measurements and flights in October 1968 and 1969, two extensive measuring sessions were held near Den Dool in the winter of 1970–71 and in the winter of 1974–75. The winter of 1974–75 was preceded by a very wet autumn, and all the grass had been eaten by a plague of mice. Table 7 gives the damping depths and the approximate phase differences measured.
Figure 29. A typical diurnal temperature cycle calculated with the soil-heat model for the test site near the river Waal. For description of the model see Appendix.

Figure 30. Diurnal temperature cycle measured at the test site near the river Waal for three different periods. Note the resemblance with Figure 29, i.e. the model simulation.
Table 7  Damping depth D (measured in the Alblasserwaard polder in the winter of 1970/71 and 1974/75) and the time difference between the temperature maxima for sand and peat. 

' + ' indicates that the temperature of the sand is ahead of that of peat, and ' − ' indicates the opposite.

The damping depth was not determined for the top layer in the winter of 1970/71.

<table>
<thead>
<tr>
<th>Time in the winter of 1970/71</th>
<th>Diurnal damping Depth D,</th>
<th>Time difference</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sand</td>
<td>Peat</td>
<td></td>
</tr>
<tr>
<td></td>
<td>deep layer</td>
<td>top layer</td>
<td>deep layer</td>
</tr>
<tr>
<td>26/2 - 2/3</td>
<td>15</td>
<td>-</td>
<td>8.5</td>
</tr>
<tr>
<td>27/3 - 1/4</td>
<td>16</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>7/4 - 12/4</td>
<td>15.5</td>
<td>-</td>
<td>8.3</td>
</tr>
</tbody>
</table>

Winter 1974/75

|                            | 13 | (6) | 10 | - | 0 to - |
| 22/12                      | 14 | (3) | 11 | (3) | 0 to - |
| 2 + 3/1                    | 20 | (4.5) | 13.5 | - | 0 |
| 8/1                        | 10 | - | (7) | 0 |
| 4-6/2                      | 11 | 2.2 | 8 | 3 | 0 |
| 8-11/2                     | 11 | 3.5 | 8 | 3.5 | 0 to + |
| 18-21/2                    | 12 | (4) | 10.5 | (3) | 0 |
| 27/2-2/3                   | 11 | - | 8 | - | 0 |
| 16/3                       | 11 | - | 7 | - | -30 |
| 21-23/3                    | 12 | - | 8.3 | - | 18 |
| 31/3                       | 11 | - | 7.5 | - | 40 |

The aim here was to detect sanded-up, fossil riverbeds that show temperature contrasts with their surroundings owing to a difference in their hydrological behaviour (see Fig. 31a). As in other measurements, values of 3.5–4.5 K were found for the temperature difference between surface water and deeper-lying water, measured in September.

These observations again indicated the importance of the vegetation cover, in this case grass. The hydrological differences gave rise to a different type of vegetation, the vegetation over seepage sites and sand veins being generally coarser. This in turn had an effect on the microclimate, and thus ultimately on the radiation temperature. From October to February the sand vein seemed to warm up more quickly than its surroundings, consisting of peat and clay. In March and April, however, an oppositely directed temperature contrast was often observed, the sand vein being colder than its surroundings. This can be fully explained by the emergence of grass, which entirely determines the radiation temperature.

3.1.4  Seepage of water from waterways

The water balance in the level stretches between the sluices in the South Willemsvaart,
Fossil river beds show temperature contrasts with their surroundings because of a difference in hydrological characteristics. Test site near Den Dool, Alblasserwaard Polder.
Figure 31.b  Situation of the fossil river beds, test site near Den Dool, Alblasserwaard Polder.
the Wessem-Nederweert, and the Noordervaart canal system (see Fig. 21) cannot be maintained fully. The sluices in this system date back to Napoleonic times and are thus liable to leak. Since the canal system lies on sandy soil, water may filter out through the bottom or the walls in some places.

The canals run in straight lines and lend themselves well to airborne thermal mapping.

To investigate these seepage phenomena, in April 1977 we used both thermal mapping and false-colour photography along a 6-km stretch of this waterway. Since the water level in the canal is constant, any seepage must remain almost uniform throughout the year. As a result, the escaping water exerts a constant effect on the vegetation in the adjacent areas. If the times of the maximum temperature contrasts are to be established, the thermal fluxes and the phase differences must be known. Since these were not known for this area, it was assumed as an approximation that the maxima occurred at 2:00-4:00 p.m. (Central European Time) (see Section 3.1.1). The effect of the moisture content of the soil on the growth of the vegetation can be seen at the beginning of the growing season (see Section 3.1.1 for the creek system in the Autriche polder, together with Fig. 22).

Fig. 32a shows the thermal map for part of the South Willemsvaart canal, and Fig. 32b a false-colour photograph for the same area. In the interpretation of the thermal images we bore in mind the isohypse map and the topography in the polder with respect to the water level in the canal. The combination of these data indicated possible spots of seepage, which were then looked at more closely with the aid of thermal mapping and false-colour photography. On the basis of these indications it is possible to choose some areas for an optimum field test along the 60-km stretch. The objectives of this application of the remote sensing method were chosen as a basis for further detailed seepage studies.

3.2 Oceanography

3.2.1 Sea surface temperature

Observation missions were flown over an area of 900 km$^2$ near the island of Texel in 1971 and 1972 to study the very complicated movements of the water (see Fig. 33) and to make visual records of the exchange of water between the Wadden Sea and the North Sea with the aid of an infrared line scanner (Brunsveld van Hulten and Kraan, 1977).

The maximum temperature difference between the Wadden Sea and the North Sea is about 3 K. The difference between the water temperature at Den Helder and at the Texel light vessel was found to be 2.1 K, partly because of the warming up that occurred in the shallow Wadden Sea during the test period as a result of insolation. Thermal mapping was therefore possible.
Figure 32a Thermal map of a part of the Zuid-Willemsvaart, taken at April 18th, 1977. Seepage areas show up lighter (warmer).

Figure 32b False colour photograph of a part of the Zuid-Willemsvaart, taken simultaneously with the thermal image. Vegetation at the seepage areas are under stress and show up light coloured.
Figure 33. Diagrams of the tidal movements at Den Helder, reconstructed from the Stroomatlas Noordzee (1963).
The use of an airborne radiation thermometer (Kraan, 1977) in the field of oceanography could be combined with thermal mapping.

The sea truth parameters were the radiation temperature, the bucket temperature and the sea water temperature measured at depths of 1 and 5 m. In 1972 the aircraft was flown at altitudes of 500 and 5000 ft to get a better idea of the atmospheric corrections. The '60° test' technique was also used in this experiment (see Section 2.4.1).

The 1971 data obtained by the airborne and shipborne infrared radiation thermometers and by the determinations of the bucket temperature were processed to obtain isotherm maps (see Fig. 34). The pictures obtained by the infrared line scanner did not contain the right information to construct a synoptic composite. Although the patterns are on the whole similar, they also show marked differences. For example, the airborne measurement differs from the sea truth at the Texel light vessel. There are also significant differences between the bucket temperatures and the data given by the shipborne IR radiation thermometer. The differences in absolute temperature can be explained by experimental errors and inaccuracies in the parameters involved in the correction factors. Differences in the patterns point to an inaccuracy in fixing the position of the aircraft and the ships, and also in the manner of transposing the shipborne measurements to the tidal phase of the flight time. The error due to subjective interpolation (0.25 K) seemed negligible, being about one-third of the standard deviation in the radiation temperature of a certain track.

The 1972 airborne measurements with an IR radiation thermometer and the measurements of the bucket temperature were converted into isotherm maps and a map of temperature fronts showing the visible temperature discontinuities in the thermal images (see Fig. 35). The airborne observations gave a detailed picture, and the fronts showed good agreement with the corresponding pattern of tidal currents (see Fig. 33) given in the 1963 'Stroomatlas Noordzee'. This agreement is much closer than in the case of the 1971 data, probably because the mixing and the displacement were much better developed in the 1972 observations. In 1971 we flew the aircraft at the turn of the tide, and in 1972 we flew it just before low tide. There was reasonable agreement between theory and experiment within the standard deviation. The difference between the corrected radiation temperature (12.9°C) and the mean bucket temperature (13.5°C) agreed well with the difference between the calculated radiation temperature and bucket temperature (− 0.7 K). The difference between the measurements at 500 ft and those at 5000 ft (Table 8) can be explained by the inaccuracy of the radiation thermometer and by the inaccuracy of the parameters used in the corrections. At that time the thermal images could not all be processed further in view of the large amount of data involved.
Figure 34. Isotherms (°C) at the North Sea near Den Helder, reconstructed from measurements of the airborne IR radiometer (top), and the shipborne IR radiometer (middle), and the bucket temperature (bottom), 1971. Depths in meters.
Figure 35. Isotherms (°C) at the North Sea near Den Helder, reconstructed from measurements of the airborne IR radiometer (top), and the shipborne bucket temperature (middle), and temperature fronts (bottom), plotted from IR line scanning images, 1972. Depths in meters.
<table>
<thead>
<tr>
<th></th>
<th>H = 500 ft</th>
<th>H = 5000 ft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean radiation temperature recorded ( (T_r) ), °C</td>
<td>13.5</td>
<td>12.0</td>
</tr>
<tr>
<td>Calculated corrections for the atmospheric influence and the reflection ( (A + R) ), K</td>
<td>+ 0.8</td>
<td>+ 1.4</td>
</tr>
<tr>
<td>Experimental correction for the atmospheric influence ( (A) ), K</td>
<td>+ 0.2</td>
<td>+ 1.5</td>
</tr>
<tr>
<td>Calculated correction for the reflection ( (R) ), K</td>
<td>+ 0.5</td>
<td>+ 0.5</td>
</tr>
<tr>
<td>Combined correction ( (A + R) ), K</td>
<td>+ 0.7</td>
<td>+ 2.0</td>
</tr>
<tr>
<td>Shutter correction ( (C) ), K</td>
<td>-0.8</td>
<td>-1.4</td>
</tr>
<tr>
<td>Skin temperature ( (T_r + A + R + C) ), °C</td>
<td>13.4</td>
<td>12.6</td>
</tr>
<tr>
<td>Skin temperature difference, K</td>
<td></td>
<td>-0.7</td>
</tr>
</tbody>
</table>

\(^{11}\) See sect. 2.3.4. for details relative to the influence of the ambient temperature

It can be concluded that the data obtained with the infrared radiation thermometer are not sufficiently detailed to give a close representation of the complicated pattern of mixing. By contrast, thermal maps are suitable for this purpose, provided that the large amount of data can be properly processed.

### 3.2.2 Large-scale circulation

After some preliminary work done in 1971 and 1972 (Brunsveld van Hulten and Kraan, 1977), and having gained the experience at Texel, in 1973 we fully tested the use of thermal mapping, in conjunction with shipborne measurements, for the study of mixing, spread, and circulation phenomena near the mouth of the river Rhine at the Hook of Holland. The aim was to monitor the river water during a complete tidal cycle of 13 hours to reconstruct its movements and mixing, and thus to produce a three-dimensional representation of the water displacement. In 1972 it was found that airborne observations at the turn of the tide, i.e. at the extreme points of the horizontal water displacement, give information only about the movement of water normal to the coast but not about its movement parallel to the coast. For the latter it was necessary to gather information more frequently over the tidal cycle.
The greatest average temperature contrast between sea water and fresh water exists in the period of May-July (see Fig. 36). The sea is then colder than the river, and this reduces the mixing between their waters. The temperature difference supplements the density difference due to a difference in salinity. In winter an opposite temperature difference prevails, which reduces the salinity-determined density differences. The influence of the temperature on the density is an order of magnitude smaller than the influence of the salinity. However, it has a noticeable effect on the mixing, so that spring seems to be better for thermal mapping.

Figure 36. Yearly mean-temperature cycles of sea water over 5 years at the harbour of Rotterdam (fresh water) and the Light Vessel Goeree station (sea water). The largest temperature contrast exists in the period May to July included. The phase shift between spring-summer and autumn-winter period is well known to exist.

An area measuring 20 × 15 km by the Hook of Holland (see Fig. 37) was systematically scanned nine times during a 13-hours tidal period (see Fig. 38), the flight time being 50 min for each observation. The constant motion of the water caused shifts between the image strips obtained during the same observation session, but these could be transposed to the centre of the observation period with the aid of the information from thermal images.

We flew a greater number of missions when the horizontal velocity of the water was at its highest (i.e. at high tide and low tide) than at the time when the tide was turning (see
The results were used to construct the water temperature distribution during the tidal cycle (see Fig. 39). The successive diagrams clearly indicate the horizontal displacements that occur during this period. The movements involved were given a vivid visual representation in speeded-up form by animation techniques.

Figure 37. The mixing area of the river Rhine water in the North Sea at Hook of Holland (situation May 1973). Covering time totals about 50 min.
Under the prevailing conditions the movement of the river water describes a round bell-shaped curve between low and high tide, and the bell-shaped area is then pushed in a north-westerly direction by the slope, the tidal current, and the wind.

![Diagram](image)

Figure 38. Distribution of the successive coverages (Fig. 37), over the tidal cycle, May 15th, 1973. Coverages are dense during HW and LW periods because of the maximal horizontal velocities and hence most rapid displacements.

Although the temperature drop caused by the mixing in the water and the heat transfer to the air is not small (3–4 K), individual areas of water remain readily distinguishable during the tidal cycle.

The salinity and the temperature were measured at depths of 2 and 5 m from two ships in the same area in the period of 14th–22nd May 1973. IRT measurements were made from one ship. To obtain enough information for the area, the ships covered the same area all the time (see Fig. 37), and a time analysis was performed on the data. It was assumed that, if the discharge volume of the Rhine and the direction and force of the wind do not vary too much during the period in question, the tidal movements are comparable cycle after cycle and averaging per tidal phase is permissible.

The measurements covered 16.25 tidal cycles, 90% of which were used. The area covered was divided into a number of grid areas for the time analysis, the size of these graticules being determined by the accuracy of the measuring instruments with regard to time, place, and magnitude.

Since we were dealing here with a large-scale stochastic phenomenon, the errors introduced by the above assumptions were of an acceptable order of magnitude in this approximation. The accuracy can be improved by using more ships for the measurements.

Since the weather was calm and stable, the hydrological boundary conditions did not change much during the shipborne measurements: the discharge of the river Rhine changed from 2770 to 2169 m³/sec, while the wind velocity varied between 8 and 23 m/sec and the wind direction between −100° and +100°.
Figure 39. Isotherms in the mixing zone at the mouth of the river Rhine, reconstructed from airborne IR radiometric measurements. Successive tidal phases are indicated (see also Fig. 38).
The data for the depths of 2 and 5 m, given by the time analysis in the same tidal phases as those involved in the thermal maps (see Fig. 38), showed reasonable correlation. This is shown in Fig. 40 for the tidal phase between HW and LW. The correlation is strongly disturbed locally by stratification and three-dimensional turbulence. The expected close correlation between temperature and salinity, resulting from the stratification, can be clearly seen from the diagrams. The correlation between the conductivity and the temperature was 0.96.

One of the conclusions we can draw from this analysis is that the ships did not follow the best course. Instead of the NE-SW direction we should have chosen NW-SE, in which direction the gradient is greatest. It can also be concluded that the temperature contrasts between sea water and river water, which show up as sharp lines, need not coincide with the isotherms.

It is thus possible to reconstruct the horizontal motion of the surface water with the aid of thermal mapping, and when animation techniques are used a further dimension is added to the interpretation of the horizontal dynamics.

The method described here can be used for the systematic investigation of mixing processes in coastal areas. The radiation temperature can be correlated with the bulk temperature, the difference between them being about 1 K (see Fig. 19). Besides, one can obtain an idea of the salinity, where there is high correlation between the conductivity and the temperature (in the spring). It is therefore possible to study how the systematic motions of water are affected by the tidal movements, the volume discharged by the river, and the wind direction. The results show the net coastal transport of soluble substances discharged into the river.

The movement of water in the coastal area is of course complicated and can only be described adequately in three dimensions. Here, thermal mapping gives additional information covering not only the horizontal movement but also the effect of the deeper layers when they surface. The latter can be studied in pictures that show up the cold water at the surface (see Fig. 39 for the area directly outside the estuary of the New Waterway). The temperature contrasts are smallest in February, March, and September-November (see Fig. 36). Mixing by storms is strongest in these months and less strong in April to August. The probability that a stationary situation will occur, with the build-up of the concentration level, is thus much greater when the temperature contrasts are highest.

3.2.3 Large-scale annual variations

Ten IR radiometric missions were flown along the Dutch coast between February 1972 and February 1975 (Kraan 1977). The area extended between the Hook of Holland and
Figure 40. Temperature and salinity distributions for the tidal phase between HW and LW, reconstructed from airborne and shipborne measurements. The layered structure in the mixing zone due to differences in fresh water and sea water, affects the correlation between the different levels. However correlation between temperature and salinity is striking.

the island of Texel (see Fig. 41), covering 2000 km². The aim was to establish whether the annual temperature cycle could be detected with sufficient accuracy over such a large area using only limited aerial observations. For such large-scale temperature distributions the highly detailed information supplied by an infrared line scanner is redundant and one can do with an infrared radiometer (see Fig. 41). The readings were reasonably stable over a period of about 3 hours, and were within the diurnal variations. The total random error in the temperature gave a standard deviation of 0.4 K.

The ten flights were flown over a period of 8 months, i.e. four monthly pictures are missing from the annual cycle. To take into account the movements of the water in the course of the flight, we also compared two isotherm maps, one with transposition and one without. This transposition gave no improvements, so that flying with the tide within 3 hours does not make transposition necessary.

It proved possible to show the distribution of the sea surface temperature using a grid measuring 1.5 x 1.5 km². The standard deviation of 0.44 K showed that the infrared radiometer provides useful results in such cases.
Figure 41. Part of the Dutch coast covered in 3 hours with an airborne IR radiometer, from which the thermal distribution over the sea surface has been constructed. Temperature in °C.
Satellite images (Fig. 42) could also be used for establishing the distribution over very large areas of slowly changing temperatures. The pattern of temperature distribution shown in this thermal map is in good agreement with the mean surface temperature for August (see Fig. 43). However, the mean cloud cover, which on average amounts to
more than 60% in our area, is very restrictive. Besides, the transit time of a satellite, which passes overhead once every 16 days at the same time of day, generally does not coincide with the right tidal phases, so that this method provides little or no information about tidal movements. Satellite pictures can therefore as a rule only be utilized for very large-scale phenomena that take place slowly, over a period ranging from a season to several years. They are unsuitable for instance for places where tidal movement is pronounced, i.e. inshore.

Figure 43. Surface temperature distribution (°C) over the North Sea, August, Monthly Means 1905-1954 (ICES 1962).
3.3 Environment

3.3.1 Cooling water discharges from power stations

In 1971 we started using the thermal mapping method to inventorize the discharge of cooling water from the most important Dutch power stations. At that time the infrared line scanner was still unsuitable for the determination of the radiation temperature in absolute terms, and gave only qualitative pictures. For subsequent work with improved equipment we chose some power stations where the water into which they discharged had certain specific hydraulic characteristics which could be detected by thermal mapping. The fact that measurements from ships cannot supply detailed information highlights the usefulness of remote sensing. The flights were of course still accompanied by in-situ measurements to estimate the atmospheric influence and the skin-layer of the water.

One of the aims here was to determine the capacity of the receiving water body, which depends on its size, the flow and the extent of mixing. Another aim was to check the temperature increase and the amount of heat discharged. Although thermal mapping only provides information about the surface, it can disclose some interesting aspects that cannot be observed by other techniques. In particular, it enables one to study in detail the circulation and mixing in the horizontal plane.

The amount of cooling water discharged varies with the amount of electricity generated by the power station. The capacity of the receiving water body depends partly on its nature, i.e. on whether it is a river, stagnant or semi-stagnant basin, a tidal river, or a canal with locks and sluices. The spread can be established reasonably well from a series of observations adjusted to variations in the parameters mentioned. For example, in rivers the cooling water does not mix well for many kilometres downstream, and this leads to higher excess temperatures over smaller areas than would be produced in the case of complete mixing across the width of the river.

3.3.1.1 The Velsen Power Station

The Velsen power station has an installed capacity of 1042 MW and is situated near the Orange Sluices on the North Sea Canal (see Fig. 21). As sea-going vessels are gated through these, salt water enters the 16 m deep Velsen basin and gives rise to stratification. The fresh water from the canal is used by the power station as cooling water, which is then discharged into the Velsen basin via the inner harbour canal. Recirculation can occur, because the inlet and the outlet lie fairly close to each other.

We used thermal mapping to study the spread of the cooling water under certain fixed conditions. When the water is stationary, this cooling water spreads over the Velsen
basin, as can be seen from Fig. 44a. When the locks are used, however, the cooling water is discharged via the outlet sluices towards the harbour of IJmuiden and spreads further over the North Sea (see Fig. 44b).

Industrial discharge east of the Velsen power station has a noticeable effect. When the locks are in use part of this discharge enters the harbour, from where the cooling water is taken. The pictures also show numerous further details concerning the movement of water e.g. around the outlet sluice and the splitting-off of some cooling water that can persist around the Orange Sluice. A great deal of detail can also be seen in the IJmuiden harbour.

Figure 44. Thermograph over the Velsen Basin and IJmuiden harbour at two different moments.
1. At that moment a mammoth tanker is sluiced with the help of two tugboats at the rear.
2 and 3. Distribution of the cooling water over the Velsen Basin and the harbour of IJmuiden. Light is warm and dark is cold.

3.3.1.2 THE FLEVO POWER STATION

The Flevo power station has an installed capacity of 835 MW and is situated at the relatively shallow IJsselmeer, the former Zuiderzee, about 4–5 m deep (see Fig. 21). The
Figure 45. The Flevo power station shown with the cooling-water streak turned around the power station island, the situation for a potential recirculation risk.
spread of the cooling water is affected by the wind-driven circulation in the semi-stagnant receiving water basin and by the wind on the surface. The vertical mixing depends mainly on waves and so ultimately on the wind.

The area where the cooling water is discharged was overflown a number of times for thermal mapping, which enabled us to establish the movements of the cooling water. In one case the cooling-water streak turned round the power station towards the southwest (see Fig. 45), with a potential risk of recirculation. In another case the cooling water formed a fairly narrow streak along the dike over a 10 km stretch, extending in the north-eastern direction to the IJssel bridge (see Fig. 46).

![Figure 46. The Flevo power station with the cooling-water streak extending to the north-east along the polder dike.](image)

The horizontal temperature distribution was established to within 0.1 K from the thermal maps. The infrared thermometer gave the absolute temperature with an accuracy of 0.5 K, with an absolute limit set by the corrections described in Sections 2.3 and 2.4. The outlet temperature was determined within the accuracy of the two instruments used (radiometer and scanner). With low-altitude flights (500 ft) the geometric resolution was sufficiently good to show the small-scale temperature distribution.

3.3.1.3 THE AMER POWER STATION AND THE MOERDIJK INDUSTRIAL AREA

The Amer power station (installed capacity 2198 MW) is situated south of the
confluence of the river Donge and the Bergsche Maas (Diverted Meuse), at the point where the latter turns into the Amer (see Fig. 21). To the north the Spijkerboor flows into the Amer. On the right bank of the Amer lies the Biesbosch nature conservation area.

The Moerdijk industrial area is downstream, past the Moerdijk bridge, on the south bank of the Holland Diep.

In earlier times the Biesbosch was directly connected with the sea via the Delta estuary. Despite the closure of the sea inlets, still some communication with the North Sea through the New Waterway is left, which greatly reduced the tide, extending up to the Amer (see Fig. 47). This tide and the water carried by the Meuse ensure a spread of the cooling water from the Amer power station and from the Moerdijk industrial area, lying about 20 km downstream.

Figure 47. The actual reduced tidal cycle and the river discharge at the Amer power station.

Four flights were made over this power station, covering a complete tidal cycle, the aim being to establish the spread pattern of the cooling water (see Fig. 48). Fig. 49 shows the absolute temperature distributions at successive times. The shallow and stagnant waters of the Biesbosch are affected by the cooling water, which enters when the tide rises, and are heated by insolation. The exchange between these basins and the Amer is slight. The contributions of the cooling water and the insolation will have to be established separately, because only the combined effect is measured.
Figure 48. Thermographs at indicated tidal phases near the Amer power station.
Fig. 49 also shows clearly the restricting effect exerted on the mixing by the guiding dam erected parallel to the flow between the Amer and Merwede. The current and the temperature were measured continuously from boats during the entire tidal cycle. Fig. 50 shows the temperature distribution at a depth of 0.5 m in almost the same tidal phase as that in Fig. 48. Comparison of the two pictures shows that thermal mapping provides more detail.

The falling tide can carry the cooling water over a long distance before it becomes completely mixed. Since this might reduce the capacity of the downstream stretch of the river, known as Holland Diep, to accept the discharge of cooling water, we included the
Figure 50. Surface measured temperature distribution (°C) at a depth of 0.5 m, corresponding to the tidal phases in Fig. 48.

Holland Diep basin in the aircraft mission (see Fig. 51). It was found that the cooling water released by the Amer power station can raise the temperature at the Moerdijk bridge by 1–1.5 K.
3.4 Miscellaneous

3.4.1 Cooling water discharges from chemical plants

The large volumes and the high temperatures of the cooling waters discharged by large chemical plants can threaten the receiving surface waters with thermal pollution. This is the situation in the port of Rotterdam (see Fig. 21), where thermal pollution takes place together with oil pollution caused by leaks, deliberate discharges from ships, and accidents during transfer operations. The radiation temperature of the water is thus affected simultaneously by two separate factors, heat and oil. To distinguish between them full-color photography was used in addition to thermal mapping and IRT. Since oil and water have different refractive indices and reflection coefficients (see Section 2.4.3), the sunlight reflected from the surface clearly shows any oil spill in a photograph. Since the entire visible region of the spectrum is involved in this type of photography, the colours also give a relative measure of the thickness of the oil layer on the surface: thin layers are bluish-grey and thick ones are brownish-yellow (see Fig. 52). Although some more advanced photographic techniques are available (Vizy, 1974), colour photography is mostly chosen, because it is simple and reliable.

Fig. 53 shows in detail the same part of the harbour of Rotterdam as Fig. 52. There is little agreement between the thermal image and the full-colour pictures. Oil is easy to locate by its different reflection coefficient in the visible region. However, the emission coefficient of oil shows such a large scatter (cf. 2.4.3) as the oil ages on water that temperature corrections cannot be readily effected. However, it is possible to detect particularly hot sources with an excess temperature of 10 K and thus to estimate approximately the area of the harbour and river affected (see Fig. 54 for the temperature distribution). The distribution of oil in the port of Rotterdam is illustrated in Fig. 55, which is derived from the full-colour pictures.
3.4.2 Oil spill detection

Oil spills and slicks can be detected by thermal mapping because oil and water have different emission coefficients. As already mentioned in Section 2.4.1, the effective emission coefficients $\bar{\varepsilon}$ for water are 0.99 and 0.98 in the 9.5–11.5 $\mu$m and the 8–14 $\mu$m regions, respectively. In contrast, the emission coefficient of oil varies widely with the type of the oil and the thickness of the oil layer (see Sect. 2.4).

Figure 52. Full-colour photograph of oil slicks over the New Meuse, taken simultaneously with thermal mapping: May 21st, 1974. The oil shows up by having a different colour from the water due to refraction of light in the layer. The layer thickness can be estimated from the shade of the colour.
Figure 55. Oil pollution in the harbour of Rotterdam as revealed by full-colour photography on May 21st, 1974.
Since the emission coefficient of sea water is quite different from that of oil, the amount of radiation recorded is difficult to correlate with the temperature of the oil layer. The radiation temperature is affected by the heat exchange between the air and the water through the oil layer, and this heat exchange is controlled completely by the oil layer. Although the process is not yet clearly understood, it can be described generally as follows. The normal evaporation of water is fully suppressed and is replaced by

Figure 56. Thermal maps of an oil spill taken during a controlled experiment at the North Sea, showing the evolution of the spill under influence of wind and tide.
evaporation of the lower-boiling oil fractions. A lower emission coefficient results in a lower radiation temperature. A restricted evaporation would result in a temperature increase. Both factors show a large scattering in magnitude. The combined effect makes it very difficult to determine the radiation temperature of the water. With the lighter oil varieties and crudes the higher-boiling, heavier, fractions are left after evaporation of the lower-boiling fractions, and when they too have evaporated still heavier ones remain. Finally only bituminous substances are left which float in lumps with a zero buoyancy. This transient aging process attributes to the complexity of the radiation temperature.

Figure 57. Photograph of an oil spill taken in the visible region during the controlled experiment at the North Sea in Fig. 56. The Vessel is the RWS M.S. Smal Agt oil combattance ship.
To establish whether oil pollution at sea can be detected by the thermal infrared method, we poured 2 m$^3$ of a light topped Iranian oil (a straight-run light fraction from distillation of the crude) on the North Sea in a controlled experiment carried out in 1975. Figs. 56 en 57 show the corresponding thermal image and an ordinary photograph for comparison. We monitored the elongated oil streak during a tidal cycle, and found that its movement and size could be traced if the correct location of the aircraft was known.

The Rijkswaterstaat took part in an experiment in which the Institut Français du Pétrole compared the following remote sensing techniques in the detection of oil spills and slicks:
- thermal infrared method in the 8–12 μm region
- multi-spectral scanning with four channels in the 0.7–1.1 μm region
- full-colour photography
- UV photography
- side-looking airborne radar (wavelength: 3 cm)

Both oil by itself and oil treated with chemicals could be detected (Bocard et al. 1976). The sea was calm, the wind was light, and there were broken clouds in the sky. Remote sensing made it possible to locate oil slicks and to study the effects of various dispersants (Finasol OSR2, Esso-Corecit 9527 and BP 1100X); Shell Oil Herder was also used to investigate its oil-containing effect. Under the conditions mentioned the combination of the thermal infrared method and photography in the visible region gave the best information.

### 3.4.3 Navigation

To adapt thermal mapping to traffic monitoring we must first decide what is to be observed and with what accuracy. The aim here is often a statistical analysis of the traffic stream density, which varies with time. The determination of the vessel classes can be considerably improved compared to standard radar techniques. Although thermal mapping is not really suitable for the identification of individual ships, it can be used to monitor ship movements by classifying them according to size from serial observations.

We first used thermal mapping to monitor shipping on 28th July 1975. Since ships are at least 10 K warmer than the sea, thermal shows them as hot spots against a cold background (see Fig. 58). By adjusting the thermal resolution and the flight level one can make out the characteristic features of ships, such as their length, the width, the forward deck, the cargo holds on deck, the bridge, and the engine room section. On this basis ships can be roughly classified reasonably well by their size and type. Records on
magnetic tape can be subjected to image processing by enhancement and enlargement for the purposes of detailed interpretation.

In the case of a sufficiently high thermal resolution we also obtain data on the temperature distribution of the water surface. The wake of a ship remains ‘visible’ for a
long period, and it is thought that the turbulent energy that the ship’s screws impart to the water is dissipated only slowly in the narrow vortex stream. This effect could also be caused by a difference in the distribution of the surfactants between the wake and the surrounding water stimulated by the turbulent movement in the vortex stream. The energy can prevent the surfactants in the skin layer of the water in the ship’s wake (see Section 2.4) from homogenizing for a long period, giving a difference in the radiation temperature. This can be used to establish the ship’s direction, even when the tide has shifted the wake. The location of the ships can be fixed with sufficient accuracy from the thermal images, since the Decca position of the aircraft platform is continuously recorded. The first application of thermal mapping to traffic monitoring in 1975 showed the position of ships in a number of shipping lanes along the Dutch coast in the form of an instantaneous picture. At the same time a second aircraft carried out ordinary reconnaissance over the same area, so that the size and the type of the vessels could be verified.

On the second occasion, in 1976, more frequent thermal mapping missions were carried out over the approach to the Port of Rotterdam, at successive times to study the movements of ships and hence the traffic stream. This is a busy place, used by vessels of many different types and sizes (see Fig. 58).

We still used ground control observations in these projects, but a better image analysis will ensure much more information from the thermal maps, so that it will also be possible to use this technique at night and without any control observations.

The results of these projects indicate that thermal mapping can provide cost-effective traffic monitoring data that are not easily supplied by other methods, if at all.

The applications described above point to new possibilities, such as studies of the vortices formed in the horizontal mixing of water masses at different temperatures.

3.4.4 Horizontal mixing

As can be seen from a thermal image of the Ketelmeer (Fig. 59), thermal mapping can be used for the study of horizontal water movements, particularly in shallow basins. The Ketelmeer connects the IJssel (a branch of the river Rhine) and the IJsselmeer (the former Zuiderzee). It receives some of the water coming from the drainage of the NW Overijssel Province via the Zwarte Meer (Black Lake). The water of the Ketelmeer is cold in December, while the water that runs into it is relatively warm. The thermal map shows clearly that the two water masses flowing into it do not mix well, mixing occurring only after about 10 km. The outlet water from the Zwarte Meer shows a vortex stream at the jet boundary. The symmetry of the individual vortices is exactly as
predicted by theory (Lamb, 1932; Hinze, 1975, Section 1.1). The way the water rotates in the individual vortices can be seen. The same picture shows the western inlet from the IJssel, though this is less clear.

It can also be seen that the jetties in Schokkerhaven, where the outflow is slight, have an important effect on circulation in the basin. Provided the altitude is not very high, the influence of the atmosphere (see Section 2.3) can be neglected. This is how Fig. 59 was obtained. The coalescing water masses are of different origins, so that the surface-active substances in them are different. The differences in the emission and reflection coefficients contribute to the contrast. A high thermal resolution is required here (1.5–3 K) to show up the mixing of the water as well. Since we are not dealing with the absolute temperature distribution, we can improve the contrast by leaving out the 8–14 μm filter. The reflection in the visible is also involved and advantage is taken of the difference in reflection coefficient because of a difference in origin of the waters. The band-width is then determined by sensor sensitivity.

3.4.5 Detection of ice

In March 1975 Rijkswaterstaat took part with a side-looking radar in the 'Sea Ice 75'
project, organized in the Gulf of Bothnia by the Winter Navigation Research Board (set up jointly by the Swedish Administration of Shipping and Navigation and the Finnish Board of Navigation) with the aim of testing various remote-sensing techniques in the visible, thermal infrared, and radar region for their ability to detect sea ice. Many different observations were made from the air over a $5 \times 5 \text{km}^2$ area, accompanied by extensive surface truth measurements. The results have been published in nine reports (see Fagerlund et al., 1976).

It was concluded that thermal mapping can be used to distinguish between water, new ice, and older, thicker ice. It gives a better contrast between new ice and ice-free water, and can be used to some extent to determine the thickness of the ice. When coupled with a few surface truth measurements, it gives an idea of the ice-thickness distribution. In the thermal images it is possible to identify some special ice features such as floes (rafting), cracks, and ridges. Aerial photography can provide further information but only during the day and in fine weather, whereas thermal mapping can also be carried out at night.
4 Conclusions

From the applications described over land and over sea, it has been found, that remote sensing by the thermal infrared method reveals small temperature differences and shows temperature distributions in a useful way. Relative data and time series of images enable to reconstruct the surface kinetics behaviour, which greatly contributes to our understanding of the phenomena observed. The infrared method reveals new aspects of the phenomena and so supplements our standard methods of survey and observation. In certain applications it has been shown to be of advantage to incorporate other remote sensing techniques such as photography, providing additional information.

As the method provides information on extended area's, for which conventional surface methods hardly can be used, it also enables us to optimize the utilization of costly field labor.

In the following applications of the infrared method it has been found contributive, although not yet completely developed as an operational tool.

In hydrology the ground-water flow near the surface and the seepage of water from a water body into the adjacent land can be detected qualitatively to a certain extend. With a thorough understanding of the heat balance and the characteristics of the soil layers, the optimal moments for maximal surface temperature contrasts can be established. To that end a heat balance model for soil is developed.

The water intrusion into the land can be studied in relation to stress situations for the crops or even potential hazards for flooding, or, to the water deficit in the water body, causing hazards to the shipping traffic.

In oceanography, the large scale surface temperature distribution was detected. The reconstruction of a particular tidal cycle at the mixing area of the river Rhine and the yearly temperature cycle over the southern North Sea have been carried out succesfully on the basis of time series of thermal scanning. The kinematics of water bodies in relation to cooling water can be studied into detail by thermal mapping. Although being in essence a three dimensional mixing problem, the horizontal temperature distribution information enables to approximate the heat load and provides guidance to optimize surface measurements.

For the three applications mentioned above, the thermal mapping method provides reasonably accurate quantitative temperature information. Oil spills, their appearance and their behaviour in spreading can be studied qualitatively by thermal mapping. The strength of the method lies in the night capability, by
which deliberate oil spill activities might be controlled in an effective way with regular surveillance. A quantitative analysis still needs further investigation into the characteristics of the emission coefficient of oil. As there are different types of oil, the thermal radiation characteristics vary widely, affected by unknown relations to their surroundings. Useful additional information on oil spills is obtained with photography, either black and white or full-colour.

A combination of oil spills and cooling water as is the case in harbours with chemical industries, turns thermal mapping useless. The radiation temperature distribution over the surface cannot be related uniquely to either that of the cooling water or that of the oil layer. Some advantage can be taken from the combined application of thermal mapping and full colour photography, but positive results may not always be expected. Ship traffic studies for navigational or statistical purposes can be helped by thermal mapping. The day and night capability provides round the clock information. Reconnaissance of the individual ships however is not possible in this way.

In particular studies, such as the mixing of two water bodies from different origin, thermal mapping can be of great use. Trajectories and the kinematics of mixing over the boundary layer, as well as vortex behaviour can be studied. The application of thermal mapping to ice detection is only mentioned.

For further applications of thermal mapping, both in the field of development as well as for operational use, quick processing and adequate presentation of relevant information is prerequisite.

Development of thorough information extraction by means of elaborate data processing is necessary for research purposes. Here improvement of the signal-to-noise ratio has been demonstrated, which is only part of it.

For full-scale operational use the overall turn-around time of the system has to be defined clearly, to guarantee the cost-effectiveness of a system. For such applications the capability of the system has to be chosen accordingly to the information requirements, such as the amount of information, and of detail, as well as accuracy and reliability. In such cases development of dedicated intelligent sensors capable of information compression may be considered, rather than having relative simple airborne sensor systems combined with ground-based powerful data treatment systems, which congests the overall transmission capability.

For more accurate quantitative temperature information, further studies on the emission coefficients have to be carried out. Each application calls for a certain accuracy with regard to the place, the temperature and the spatial resolution. Therefore further development of heat-flux models through surfaces, such as the air-water interface, will greatly support the effective use of the thermal mapping method and vice versa.
The soil heat model, based on the considerations discussed in Section 2.5.2, is shown schematically in Fig. 10. The lateral heat fluxes in the ground and atmosphere are neglected. The extent to which heat penetrates a layer or a body by conduction is determined by its thermal conductivity $k_s$ and the specific heat $c_{p,s}$. The measure of this thermal penetration is the thermal diffusivity $a_s = (k_s/\rho_s c_{p,s})$. For a periodic phenomenon we can use the damping depth $D$, defined as $\sqrt{2a_s/\omega}$, where $\omega$ is the angular frequency of a surface temperature cycle that shows an approximately sinusoidal variation with time. Since the phenomena under consideration depend on the position of the sun, we use the true solar time. In a homogeneous, isotropic soil (Van Wijk, 1966), the temperature can be approximately described by the following expression:

$$T_s(z,t) = T_s(0,0) + A e^{\frac{z}{D}} \sin (\omega t + \frac{\pi}{D})$$

where the depth $z$ is taken positive in the upward direction. As $z$ increases the amplitude decreases and a phase shift occurs, so that the peak appears later and later. As most soils are neither homogeneous nor isotropic, Eq. (1) is not entirely valid. Besides, the periodic variation of the surface temperature with time is not purely sinusoidal (see Sect. 2.5.2, Fig. 10). The periodic function can be written as a Fourier series:

$$T_s(z,t) = \frac{a_0}{2} + e^{\frac{z}{D}} \sum_{n=1}^{\infty} \left\{ a_n \cos (n\omega t + \frac{\pi}{D}) + b_n \sin (n\omega t + \frac{\pi}{D}) \right\}$$

Comparing Eqs. (1) and (2), we obtain the following result for the first three terms:

$$a_0 = 2T_s(0,0)$$
$$a_1 = 0$$
$$b_1 = A$$

The higher harmonics follow from corresponding solutions of the resulting differential equations. However, here we shall consider only the first harmonic $A_1$ for $A$, and the fundamental tone $\omega_1$ for $\omega$.

The surface is exposed to turbulent air currents. During the day the air is heated by radiation emitted by the soil, and the heat is distributed vertically over a thick layer of air. If we assume vertical heat transport downwards, which is analogous to thermal...
conduction, then the apparent thermal conductivity must vary with the height, i.e. $k_s = k_s(z)$. The efficiency of the heat transfer depends on the intensity of the air turbulence. The latter increases greatly with the wind, but also with atmospheric instability, in which case the buoyancy force contributes to the heat transport.

To construct a 'heat model' we start from the schematic representation shown in Fig. A1. We assume that the evaporation ($Q_{La}$) is slight (which is certainly true during winter), and the phases of the sensible-heat flux in air ($Q_{Ha}$) and in the ground ($Q_{Hs}$) are not very sensitive to changes in $Q_{Hs}$, while the surface temperature is. This means that the variations in the heat-balance terms inside the layer $(0, z_o)$:

$$Q^* + Q_{Hs} + Q_{Ha} + Q_{ir} = 0$$  \hspace{1cm} (3)

are essentially accounted for by the radiation emitted by the ground ($Q_{ir}$), i.e. by the surface temperature. The phase shift in $(Q_{Ha} + Q_{Hs})$ caused by differences in the thermal properties of the ground can then be transformed into a phase shift in the surface temperature.
Heat flux in soil

In the case of a homogeneous and isotropic soil, the heat flux is described by the following equation:

\[ \frac{\partial T_s(z, t)}{\partial t} = \frac{\partial}{\partial z} \left\{ -a_s \frac{\partial T_s(z, t)}{\partial z} \right\}, \quad -\infty < z \leq 0 \]  

(4)

where

\[ a_s = \frac{k_s}{\rho_s c_{p,s}} \]

For \( z = -\infty \), \( T_s(-\infty, t) = T_s(0, 0) \), and for \( z = -z_0 \) we have the following expression (see Eq. 1):

\[ T_s(o, t) = T_s(0, 0) + A e^{i \omega_1 t} \]  

(5)

Let us assume that the solution of Eq. (4) can be written in the form:

\[ T_s(z, t) = \xi_s(z) \cdot \psi_s(t) \]

(see Chapter IV in Watson’s book, 1966). The solution of Eq. (4) is then as follows:

\[ T_s(z, t) = T_s(0, 0) + A e^{i \omega_1} e^{i \omega_1 t + \varphi_1} \]  

(6)

where

\[ D_1 = \sqrt{\frac{2a_s}{\omega_1}} = \sqrt{\frac{2k_s}{\rho_s c_{p,s} \omega_1}} \]

Eq. (7) gives the variations in the temperature and the thermal flux densities in a two-layer system for the soil (Van Wijk, 1966), in the case of the fundamental tone \( \omega_1 \), at the surface \( z = 0 \):

\[ T_s(o, t) = T_s(0, 0) + A_{1,1} \sin(\omega_1 t + \varphi_1) + A_{1,2} \sin(\omega_1 t + \varphi_2 - \frac{d}{D_2} \frac{d}{D_1}) \]  

(7)

where:

\( A_{1,1} = \) first harmonic of the incident wave
\( A_{1,2} = \) first harmonic of the reflected wave
\( \varphi_1 = \) phase shift of the incident wave
\( \varphi_2 = \) phase shift of the reflected wave
\[ d \] = thickness of the top layer
\[ D_1 \] = damping depth of the top layer

The subscripts 1 and 2 refer to the top layer and the deeper layer, respectively. Furthermore is:

\[ \varphi_1 = \arctan \frac{re^{-\sigma_1 d} \sin \left( \frac{2d}{D_1} \right)}{1 + re^{-\sigma_1 d} \cos \left( \frac{2d}{D_1} \right)} \]
\[ \varphi_2 = \varphi_1 - \left( \frac{2d}{D_1} \right) \]
\[ r = \frac{\sqrt{k_{s1}c_{p,s,1}} - \sqrt{k_{s2}c_{p,s,2}}}{\sqrt{k_{s1}c_{p,s,1}} + \sqrt{k_{s2}c_{p,s,2}}} \]
\[ \hat{A}_{1,1} = \hat{A}_1 \left( 1 + r^2 e^{-\sigma_1 d} + 2re^{-\sigma_1 d} \cos \left( \frac{2d}{D_1} \right) \right)^{\frac{1}{2}} \]
\[ \hat{A}_{1,2} = \hat{A}_{1,1} e^{-\sigma_1 d} \]
\[ D_1 = \frac{2k_{s1}}{\rho_{s1}c_{p,s,1}c_{v,s,1}} \]

We transform the interval \((0, z_{max})\) into \((z_0, z_{max} - z_0)\) by \(z' = z - z_0\).

**Heat flux in air**

The heat flux in air is given by:

\[ \frac{\partial}{\partial t} T_a(z', t) = \frac{\partial}{\partial z'} \left( a_a(z') \frac{\partial T_a(z', t)}{\partial z'} \right) \] (8)

where for \(0 < z' < (z_{max} - z_0)\) we have that

\[ a_a(z') = \kappa \cdot u_* \cdot (z - z_0) \]

For \(z = +\infty\), \(T_a(\infty, t) = T_s(o, o)\) and for \(z = z_0\), we have the following expression (see Eq. 1):

\[ T_a(o, t) = T_s(o, o) + Ae^{ko_1 t} \] (9)
with

$$\omega_1 = \frac{2\pi}{86400} \text{ (s}^{-1}\text{)}$$

The boundary conditions (5) and (9) indicate that we assume that the vegetation layer $-0 < z < z_o$ has a zero heat capacity and infinite thermal conductivity.

For $z = z_{max}$,

$$Q_{Ha}\bigg|_{z = z_{max}} = 0$$

because no heat is exchanged with the air layers above it. It follows from this that:

$$\frac{\partial T_a(z', t)}{\partial t}\bigg|_{z' = z_{max} - z_o} = 0$$

(10)

**Solution for the equation for the heat flux in air**

Let us assume again that the solution of Eq. (7) can be written as:

$$T_a(z', t) = \zeta(z') \cdot \psi(t)$$

A basic solution for $\psi(t)$ is $e^{i\omega_1 t}$. The differential equation (4) is combined with $\psi(t) = e^{i\omega_1 t}$ to give:

$$z' \frac{d^2 \zeta}{dz'^2} + \frac{d \zeta}{dz'} - \frac{i\omega_1}{a_a} \zeta = 0$$

(11)

Transformation with

$$x' = \sqrt{\frac{4\omega_1 z'}{a_a}} \quad \text{and} \quad x = x' \sqrt{i}$$

gives the modified zero-order Bessel equation:

$$x^2 \frac{d^2 \zeta(x)}{dx^2} + x \frac{d \zeta(x)}{dx} - x^2 \zeta(x) = 0$$

for arg $x = \pi/4$. According to Watson (see Section 3.7 in his book, 1966), the general solution of this equation is:
\[ \zeta(x) = AI_o(x'i^2) + BK_o(x'i^2) \]

where \( x' \) is real, and \( A \) and \( B \) are complex numbers. Transformation into \((z - z_o - \delta)\) gives the solution:

\[ \zeta(z') = AI_o(i^24\omega_1(z - z_o - \delta)) + BK_o(i^24\omega_1(z - z_o - \delta)) \] (12)

The general solution of Eq. (8) is then as follows:

\[ T_a(z',t) = [AI_o(i^24\omega_1(z - z_o - \delta)) + BK_o(i^24\omega_1(z - z_o - \delta))]e^{i\omega_1 t} \] (13)

For the boundary condition \( z' = z_{max} - z_o - \delta \) we need \( \frac{\partial T_a(z',t)}{\partial t} \). Let \( A = a + bi \) and \( B = c + di \), in which case we can write \( \zeta(x) \) in the form of Thomson's functions (see Watson, 1966, Chapter 3.8):

\[ \zeta(x) = A\{ber(x) + i bei(x)\} + B\{ker(x) + i kei(x)\} \] (14)

Differentiation of \( \zeta(x) \) with respect to \( z \) gives the following expression with the aid of the chain rule:

\[ \frac{d\zeta(x)}{dz} = [A\left(\frac{d}{dx}I_o(x'i^2)\right) + B\left(\frac{d}{dx}K_o(x'i^2)\right)] \cdot \frac{dx}{dz} \]

Writing this in the same form as Eq. (14), we obtain:

\[ \frac{d\zeta(x)}{dz} = [A\{ber'(x) + i bei'(x)\} + \]

\[ B\{ker'(x) + i kei'(x)\}] \cdot \frac{\omega_1}{a_o(z - z_o - \delta)} \]

For the boundary value \( z' = z_{max} - z_o \) we put \( x = x_L \). Using Eq. (10), we obtain Eq. (15) with the complex representation of \( A \) and \( B \), after separating the real and the imaginary part:

\[ a ber'(x_L) - b bei'(x_L) + c ker'(x_L) - d kei'(x_L) = 0 \]

\[ a bei'(x_L) + b ber'(x_L) + c kei'(x_L) + d ker'(x_L) = 0 \] (16)

The boundary condition at the interface \( z' = z_o + \delta \) gives
\[ x_o = \sqrt{\frac{4\omega_1 \delta}{a_o}} \sim o(10^{-2}) \]

According to Watson (see Section 3.8 of his book, 1966), we therefore have that:

\[ \text{ber}(x_o) \sim 1, \text{ker}(x_o) \sim 0.1159 - \ln x_o \]
\[ \text{bei}(x_o) \sim 0, \text{kei}(x_o) \sim -\frac{\pi}{4} \]

Combining these values with Eqs. (6) and (14), we obtain the solution in the form:

\[ T_s(o, t) = T_s(o, 0) + \hat{A}e^{k_o t} \]

where

\[ \hat{A} = (a + c \text{ker}(x_o) + \frac{\pi}{4} d) + i(b - \frac{\pi}{4} c + d \text{ker}(x_o)) \]

Since \( \hat{A} \) is real, it follows that

\[ a + c \text{ker}(x_o) + \frac{\pi}{4} d = \hat{A} \]

and

\[ b - \frac{\pi}{4} c + d \text{ker}(x_o) = 0 \quad (17) \]

The equations given in (16) and (17) form four first-order equations in \( a, b, c, \) and \( d, \) with the aid of which the general equation (13) can be solved.

The values of \( \hat{A}, z_{max}, \) and \( a_s \) must be determined experimentally. \( \hat{A} \) and \( z_{max} \) are strongly time-dependent, while \( a_s \) varies with time to a much smaller extent.

This approximation has been tested against field measurements (see Section 3.1). For the higher harmonics we have the same equations, with the same solutions, which can be superposed. The model can be refined by using a Fourier expansion for the upward heat flux.

Model simulation

First the basic harmonic \( Q_{hs} \) of the sensible heat flux in the clay layer is determined. This layer is composed of a top layer of 10 mm thick and a nearly homogeneous half infinite sublayer. The data measured are in the following table A.
Table A. Soil data of the test-site near the Waal

<table>
<thead>
<tr>
<th>Parameter</th>
<th>top layer</th>
<th>sublayer</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho C_{p,s} \text{ (WK}^{-1}\text{m}^{-3}) )</td>
<td>3.1 ( \times ) 10^6</td>
<td>2.5 ( \times ) 10^6</td>
</tr>
<tr>
<td>( k_s \text{ (WK}^{-1}\text{m}^{-1}) )</td>
<td>7.10^{-2}</td>
<td>1.40</td>
</tr>
</tbody>
</table>

From the surface temperature variation measured the fourier component of the basic and the first harmonic together with their phase have been established. Results are summarized in table B.

Table B. Measured surface temperature amplitude and phase.

<table>
<thead>
<tr>
<th>Basic harmonic</th>
<th>First harmonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude (K)</td>
<td>Phase (rad)</td>
</tr>
<tr>
<td>2.66</td>
<td>-2.3</td>
</tr>
</tbody>
</table>

These data enables to calculate the temperature variation in the various layers. Furthermore the sensible heat flux \( Q_{Hs} \) can be established by means of the general relation

\[
Q_{Hs} = -\rho_s C_{p,s} a_s \frac{\delta T}{\delta z} |_{z=0}
\]

For the case considered the following has been calculated:

\[
Q_{Hs} = 10.5 e^{i(\omega t + 0.783)} + 7.9 e^{i(\omega t + 1.283)}
\]

Next the sensible-heat flux \( Q_{Ha} \) in air is determined. From the data the values in Table C are obtained.

Table C. Parameter values for the calculation of the sensible-heat flux.

<table>
<thead>
<tr>
<th>( u_*(\text{m} s^{-1}) )</th>
<th>( k(\cdot) )</th>
<th>( L(\text{m}) )</th>
<th>( z_o(\text{m}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.38</td>
<td>150</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Combination of the parameter values of Tables B and C enables to establish the temperature profile in air.
With the relation
\[ Q_{H_{a}} = -\varrho_{a}c_{p,a}a_{a} \frac{\partial T}{\partial z} \]
where
\[ a_{a} = \kappa u_{*}(z - z_{o}), \]
the sensible-heat transport is found. The following result is obtained.
\[ Q_{H_{a}} = 39.9e^{j(\sigma t + 0.501)} + 23.7e^{j(\tau t + 0.200)} \]
Finally the sensible-heat flux \( Q_{H} \) is found by summation of the sensible-heat fluxes for soil and air. For the case considered, the following results:
\[ Q_{H} = Q_{H_{a}} + Q_{H_{s}} = 48.2e^{j(\sigma t + 2.911)} + 27.9e^{j(\tau t - 0.414)} \]
In Fig. 30 (see Section 3.1.2), this expression is visualized.
References


ICES, 1962, Mean monthly temperature and salinity of the surface layer of the North Sea and adjacent waters from 1905 to 1954, International Council for the Exploration of the Sea, Hydrographic Service, Charlottelund Slot, Denmark.


LAMB, 1932, Hydrodynamics, Cambridge.


VAN DEN ABEELE, G., JACOBS, A. and PALLAND, C., Measurements of sensible heat transfer in the surface layer as part of an automatic station for hydrological purposes, Techn. Conf. Automated Weather Stations, University of Reading, United Kingdom.


In the series of Rijkswaterstaat Communications the following numbers have been previously published:

No. 1.* Tidal Computations in Shallow Water  
Dr. J. J. Dronkers† and prof. dr. ir. J. C. Schönfeld  
Report on Hydrostatic Leveling across the Westerschelde  
Ir. A. Waalewijn, 1959

No. 2.* Computations of the Decca Pattern for the Netherlands Delta Works  
Ir. H. Ph. van der Schaaft† and P. Vetterli, Ing. Dipl. E. T. H., 1960

No. 3.* The Aging of Asphaltic Bitumen  

No. 4.* Mud Distribution and Land Reclamation in the Eastern Wadden Shallows  
Dr. L. F. Kamps†, 1962

No. 5.* Modern Construction of Wing-Gates  
Ir. J. C. le Nobel, 1964

No. 6.* A Structure Plan for the Southern IJsselmeerpolders  
Board of the Zuyder Zee Works, 1964

No. 7.* The Use of Explosives for Clearing Ice  
Ir. J. van der Kley, 1965

No. 8.* The Design and Construction of the Van Brienenoord Bridge across the River Nieuwe Maas  
Ir. W. J. van der Ebt†, 1968

No. 9.* Electronic Computation of Water Levels in Rivers during High Discharges  
Section River Studies. Directie Bovenrivieren of Rijkswaterstaat, 1969

No. 10.* The Canalization of the Lower Rhine  
Ir. A. C. de Gaay and ir. P. Blokland†, 1970

No. 11.* The Haringvliet Sluices  
Ir. H. A. Ferguson, ir. P. Blokland† and ir. drs. H. Kuiper, 1970

No. 12.* The Application of Piecewise Polynomials to Problems of Curve and Surface Approximation  
Dr. Kurt Kubik, 1971

No. 13.* Systems for Automatic Computation and Plotting of Position Patterns  
Ir. H. Ph. van der Schaaft†, 1972

No. 14.* The Realization and Function of the Northern Basin of the Delta Project  
Deltadienst of Rijkswaterstaat, 1973

No. 15.* Physical-Engineering Model of Reinforced Concrete Frames in Compression  
Ir. J. Blauwendraad, 1973

No. 16* Navigation Locks for Push Tows  
Ir. C. Kooman, 1973

* out of print
In the series of Rijkswaterstaat Communications the following numbers have been published before (continuation)

No. 17.* Pneumatic Barriers to reduce Salt Intrusion through Locks
Dr. ir. G. Abraham, ir. P. van der Burg and ir. P. de Vos, 1973

No. 18.* Experiences with Mathematical Models used for Water Quality and Water Problems
Ir. J. Voogt and dr. ir. C. B. Vreugdenhil, 1974

No. 19.* Sand Stabilization and Dune Building
Dr. M. J. Adriani and dr. J. H. J. Terwindt, 1974

No. 20.* The Road-Picture as a Touchstone for the three dimensional Design of Roads
Ir. J. F. Springer and ir. K. E. Huizinga (also in German), 1975

No. 21.* Push Tows in Canals
Ir. J. Koster, 1975

No. 22.* Lock Capacity and Traffic Resistance of Locks
Ir. C. Kooman and P. A. de Bruijn, 1975

No. 23.* Computer Calculations of a Complex Steel Bridge verified by Model Investigations
Ir. Th. H. Kayser and ir. J. Brinkhorst, 1975

No. 24.* The Kreekrak Locks on the Scheldt-Rhine Connection
Ir. P. A. Kolkman and ir. J. C. Slagter, 1976

No. 25.* Motorway Tunnels built by the Immersed Tube Method
Ir. A. Glerum, ir. B. P. Rigter, ir. W. D. Eysink and W. F. Heins, 1976

No. 26.* Salt Distribution in Estuaries
Rijkswaterstaat, Delft University of Technology, Delft Hydraulics Laboratory, 1976

No. 27.* Asphalt Revetment of Dyke Slopes
Committee on the Compaction of Asphalt Revetments of Dyke Slopes, 1977

No. 28.* Calculation Methods for Two-dimensional Groundwater Flow
Dr. ir. P. van der Veer, 1978

No. 29.* Ten years of quality control in road construction in the Netherlands
Ir. C. van de Fliert and ir. H. Schram, 1979

No. 30. Digital Large Scale Restitution and Map Compilation
J. G. van der Kraan, H. Rietveld, M. Tienstra and W. J. H. IJzereef

No. 31. Policy Analysis for the National Water Management of the Netherlands
(Netherlands Contributions, related to the PAWN-study, for the ECE-seminar on Economic Instruments for the Rational Utilization of Water Resources – Veldhoven, Netherlands – 1980)

No. 32. Changing Estuaries
Dr. H. L. F. Saeijs

No. 33. A bird's eye view of the shipping traffic on the north sea
North sea directorate

* out of print
ten years of quality control in road construction in the netherlands

by

ir. C. van de Fliert

and

ir. H. Schram

no. 29/1979
RIJKSWATERSTAAT COMMUNICATIONS

TEN YEARS OF QUALITY CONTROL IN ROAD CONSTRUCTION IN THE NETHERLANDS

by
Ir. C. VAN DE FLIERT
State Road Laboratory, Rijkswaterstaat

and
Ir. H. SCHRAM
State Road Laboratory, Rijkswaterstaat

Government Publishing Office — The Hague 1979
All correspondence should be addressed to

RIJKSWATERSTAAT
DIRECTIE WATERHUISHOUDING EN WATERBEWEGING
P.O.B. 20907
2500 EX THE HAGUE – THE NETHERLANDS

The views in this article are the authors' own.

ISBN 90 12 02410 2
Contents

Page

5  1  Introduction
7  2  Non-statistical system (1968-1978)
14  3  Statistical system
32  4  Relationship between the penalty system and the necessary compensation costs
55  5  Summary
57  Literature
1 Introduction

Historical background

After ten years experience – involving more than 300 projects – of the system of quality control introduced in 1968 and associated with reduced-payment clauses, the State Road Laboratory has more recently developed an almost entirely statistical method.

The system which still exists at present is described in a detailed publication by Van de Fliert and Brouwers in the July-August 1968 issue of 'Wegen'[1]. That article also outlines the historical background in this area. Developments since 1960 can be briefly summarized as follows.

In the years 1960 to 1965 the first steps were taken towards the introduction of a general method of quality control for the carriageway pavements laid in a number of large highway projects; this ultimately led to an appropriate system which was accepted both by the public authorities and by the contractors' organizations.

On the basis of the large volume of data amassed over the years, contract specifications involving reduced-payment clauses were experimentally introduced for some projects in 1966. The clauses and specifications were completed in the next two years. The final version was then adopted in 1968.

At the time these developments were encouraged by a lack of sufficiently qualified, supervisory staff in the directorate and, above all, by the increasing mechanization and automation of construction work together with the growth in the scale of most projects.

Both the public authorities and the contractors gradually reached the conclusion that it was no longer acceptable to apply unilateral quality control under the sole responsibility of the directorate and normally based on analysis of a relatively small number of samples generally taken by a selective method. This method was found to be outdated and in need of fundamental change.

The clauses incorporated in the contract specifications since 1968 were adjusted in points of detail in 1972 and 1975 following a review of the general criteria and test methods for materials, mixes and surfaces. They have thus undergone no basic change since 1968. However, in recent years an alternative system has been developed, based on statistical principles and on the extensive experience of sample-taking and testing acquired over the past ten years.
Consultation between the directorate and contractors

Before the definitive system of quality control with reduced-payment clauses was introduced in 1968, consultation took place between the Public Works Department (State Road Laboratory) and a committee of expert representatives of the contractors. These consultations were intensified in the seventies when the 'Specifications for construction and control of road pavements' (VUCW) and the detailed 'Recommendations for production control in road building' (ABCW) were compiled; these texts were published in 1975. The VUCW were revised in 1977-78, again in close consultation between the public authorities and contractors' organizations, and published as the VUCW 1978.

A favourable situation obtains in the Netherlands in that the State Road Laboratory (RWL) is a central, public body with responsibility, in the road building sector, for compiling the criteria for materials and working methods, for the structural design of highways and the composition of the mixes used, for quality control specifications and for the performance of quality control in the case of state highways. This central role has led to a high degree of uniformity in the specifications applicable in the Netherlands and also to effective consultation with the representative committee of the contractors' organizations.

The system of quality control which has been applied in the past ten years has been accepted in broad outline by the Netherlands contractors' organization and recognized as reasonable, equitable and effective in maintaining the desired quality level.
2 Non-statistical system (1968-1978)

General principles

The system is based on a clear distinction between daily production control under the responsibility of the contractor on the one hand, and limited acceptance control by the directorate on completion of the works on the other. To ensure good mix characteristics, the contract specification stipulates that the contractor must effect preliminary studies to determine the thickness (and the requisite cement content) of the sand-cement roadbase, and also the composition and Marshall stability of the asphalt mixes (mix design).

The results of these preliminary studies are compared with the results of similar investigations conducted by RWL. In this way specific agreements are reached between the directorate and the contractor on the design and characteristics of the mixes before the actual work begins. In principle the method of implementation is thus the main determining factor as regards the standard of quality of the pavement layers.

Production control by the contractor

The specification stipulates that the contractor must effect thorough daily controls of the composition and characteristics of the mixes; these controls are governed by the same provisions as the acceptance control. The contractor must therefore have a well-equipped site laboratory at his disposal with qualified personnel. For major works in particular, the contractors regularly use statistical methods to control production quality; control cards are utilized for this purpose. On the basis of the data obtained in this way it is possible to control the procedures for mixing, compacting and processing the road building materials.

Acceptance control by the directorate

Quality control in respect of road building in the Netherlands relates in the first instance to the following main characteristics:
- layer thickness;
- bitumen content of the various types of asphaltic concrete;
void content of asphaltic concrete and degree of compaction of sand asphalt;
compressive strength of sand cement.

In the past 15 years no cement concrete pavements have been laid on trunk roads in the Netherlands road network. Should pavements of this kind be applied in the future a system of quality control similar to that now used for asphalt pavements would have to be developed for concrete road surfaces.

Under the system in use since 1968 one sample is taken per 2000 m² asphalt pavement; this sample consists of two cores with a diameter of about 100 mm drilled out of the completed pavement. If analysis of the characteristics listed above shows the quality

### Table 1  Table of penalties*

<table>
<thead>
<tr>
<th>layer thickness</th>
<th>penalties per 2000 m² – in guilders</th>
</tr>
</thead>
<tbody>
<tr>
<td>shortfall on thickness</td>
<td>roadbase: 0.15 m sand cement or 0.12 m sand asphalt</td>
</tr>
<tr>
<td>1- 5 mm</td>
<td>–</td>
</tr>
<tr>
<td>6-10 mm</td>
<td>–</td>
</tr>
<tr>
<td>11-15 mm</td>
<td>–</td>
</tr>
<tr>
<td>16-20 mm</td>
<td>1000</td>
</tr>
<tr>
<td>21-25 mm</td>
<td>2000</td>
</tr>
<tr>
<td>etc.</td>
<td>etc.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>compressive strength: sand cement</th>
<th>bitumen content: asphaltic concrete</th>
<th>voids: asphaltic concrete</th>
</tr>
</thead>
<tbody>
<tr>
<td>strength too low**</td>
<td>penalty per 2000 m² in guilders</td>
<td>bitumen content too low</td>
</tr>
<tr>
<td>0.1-0.5MN/m²</td>
<td>1000</td>
<td>0.1-0.2%</td>
</tr>
<tr>
<td>0.6-1.0MN/m²</td>
<td>2000</td>
<td>0.3-0.4%</td>
</tr>
<tr>
<td>1.1-1.5MN/m²</td>
<td>3000</td>
<td>0.5-0.6%</td>
</tr>
<tr>
<td>1.6-2.0MN/m²</td>
<td>4000</td>
<td>0.7-0.8%</td>
</tr>
<tr>
<td>2.1-2.5MN/m²</td>
<td>5000</td>
<td>0.8-0.9%</td>
</tr>
<tr>
<td>etc.</td>
<td>etc.</td>
<td>etc.</td>
</tr>
</tbody>
</table>

* The amounts quoted in this table were applicable for the years 1974-1977. At the beginning of 1978 most of the penalties were increased by 50%.

** Strength shortfall with reference to the criterion of at least 2.0 MN/m² applicable for the period 1972-1977. At the beginning of 1978 this requirement was reduced to 1.5 MN/m².
Figure 1. Cores, drilled out of the pavement, intended for testing in relation to acceptance control of pavements.
to be inadequate, financial penalties are imposed. These penalties are determined by the mean result for the two core samples. As a function of the gravity of the deviation from the required values, the flat-rate penalty ranges in practice from 1,000 to 10,000 guilders per sample of 2000 m² and per characteristic of a given layer, e.g. compressive strength of the sand cement roadbase or voids of the asphaltic concrete surfacing. (See Table 1.)

The system employed up to now is clearly not a genuine statistical system in the strict sense of the term since it is based on determination of the quality of individual samples and on penalties fixed in the light of the analysis results. In practice, however the number of samples is normally so large (e.g. 50 for a controlled surface area of 100,000 m²) that we have in effect a non-selective random sample capable of giving sufficient information on the quality of the work in its entirety. The results of analysis and their statistical processing lead to the same conclusion: in most cases the number of insatisfactory samples—i.e. in excess of the penalty limit—expressed as a percentage of the total number of samples is roughly the same as the number calculated theoretically from the mean and the standard deviation. In addition, the system implies that 2% of the total number of samples may show results which fail to meet the specified criteria without giving rise to penalties. On the other hand higher percentage deviations do automatically result in penalties.

In this context it is important to note that the conclusion as to whether the work as a whole is 'good' or 'bad' does not depend on just one characteristic of one component part of the road construction; each project is assessed in the light of the test results for at least 3 or 4 characteristics:

- layer thickness, strength of the sand cement (where this is used), and the density and bitumen content of asphalt mixes. In addition the tests always relate to at least 3 or 4 different layers: sand cement (15-40 cm) or sand asphalt (10-12), bitumen-bound gravel (12-24 cm), open-textured asphaltic concrete (4-8 cm) and dense asphaltic concrete (4 cm) — see Table 2.

To sum up, the overall system is thus a combination of in general some 10 different sub-systems of quality control.

In a sense the 'risks' are thus spread over the entire construction. For example, if a penalty of 1% must be imposed because of insufficient strength of the sand cement while no other penalties are charged in respect of the other characteristics and layers, the total penalty will be limited to about 0.2% of the overall value of the construction project. If on the other hand the overall penalty for a particular project is high, it may safely be concluded that the quality standard of the project as a whole is low.

Acceptance control definitely does not make routine site supervision by the directorate superfluous — quite the contrary.

Acceptance control (by RWL) on completion of the works relates only to certain spe-
cific (though essential) aspects of the construction work which are in any case only
tested on a random sample basis. Careful supervision by the local directorate (which
also supervises and takes regular note of the results of the contractor’s quality control)
can reveal, or better still prevent, extreme – and also incidental – faults. If the results
are unsatisfactory the contractor himself must take direct action and the directorate
will also ask for shortcomings to be remedied.

Results

The principal results obtained in quality control of more than 300 projects since
1968 are summarized in Tables 2 and 3.

These projects involved areas of at least 50,000 m², generally between 100,000 and

Table 2 Test results

<table>
<thead>
<tr>
<th>property</th>
<th>material</th>
<th>overall mean value $x$ ($\sim \mu$)</th>
<th>mean standard deviation $s$ ($\sim \sigma$)</th>
<th>specification or penalty limit $R$</th>
<th>$\xi$</th>
<th>$\delta$</th>
<th>$Q_{\text{calc}}$</th>
<th>$Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>compressive strength</td>
<td>sand cement</td>
<td>6.0</td>
<td>2.3</td>
<td>2.0</td>
<td>1.74</td>
<td>4.0</td>
<td>1.36</td>
<td>1.40</td>
</tr>
<tr>
<td>relative density (Marshall test), %</td>
<td>sand asphalt</td>
<td>98.0</td>
<td>2.0</td>
<td>94.5</td>
<td>1.75</td>
<td>4.0</td>
<td>1.36</td>
<td>1.40</td>
</tr>
<tr>
<td>voids, % ($V/V_0$)</td>
<td>bitumen-bound gravel</td>
<td>5.9</td>
<td>1.8</td>
<td>.95</td>
<td>2.00</td>
<td>2.3</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>open-textured asphaltic concrete</td>
<td>4.7</td>
<td>1.9</td>
<td>8.5</td>
<td>2.00</td>
<td>2.3</td>
<td>1.60</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>dense asphaltic concrete</td>
<td>3.7</td>
<td>1.65</td>
<td>7.0</td>
<td>2.00</td>
<td>2.3</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td>bitumen content, % ($m/m$)</td>
<td>bitumen-bound gravel</td>
<td>5.0</td>
<td>0.32</td>
<td>5.0 $\pm$ 0.75</td>
<td>2.34</td>
<td>1.0</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>open-textured asphaltic concrete</td>
<td>5.5</td>
<td>0.31</td>
<td>5.5 $\pm$ 0.75</td>
<td>2.34</td>
<td>1.0</td>
<td>1.65</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>dense asphaltic concrete</td>
<td>6.5</td>
<td>0.29</td>
<td>6.5 $\pm$ 0.75</td>
<td>2.24</td>
<td>1.25</td>
<td>1.57</td>
<td>6.5 - 0.65</td>
</tr>
<tr>
<td>property</td>
<td>material</td>
<td>overall mean value $\bar{x}$ ($\sim \mu$)</td>
<td>mean standard deviation $s$ ($\sim \sigma$)</td>
<td>specification or penalty limit $R$</td>
<td>$\xi$</td>
<td>quality number $Q_{\text{calc}}$</td>
<td>quality number $Q$</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------</td>
<td>------------------------------------------</td>
<td>---------------------------------------------</td>
<td>-----------------------------------</td>
<td>-------</td>
<td>-------------------------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>layer thickness, mm</td>
<td>sand cement</td>
<td>150</td>
<td>17</td>
<td>120</td>
<td>1.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sand asphalt</td>
<td>120 (130)</td>
<td>18</td>
<td>100</td>
<td>1.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>open-textured asphaltic concrete</td>
<td>40 (43)</td>
<td>8</td>
<td>30</td>
<td>1.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dense asphaltic concrete</td>
<td>40 (43)</td>
<td>6</td>
<td>33</td>
<td>1.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>total asphaltic concrete</td>
<td>200 (216)</td>
<td>20</td>
<td>180</td>
<td>1.40</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The layer thicknesses indicated in the 3rd column are nominal values, specified as minimum thicknesses. In the case of asphalt mixes, prescribed quantities must be processed: 20 kg/m² and 25 kg/m² per 10 mm nominal thickness for sand asphalt and all types of asphaltic concrete respectively. The processed quantities are measured by weighbridge and charged up to this maximum. Since the normal mean degrees of compaction of sand asphalt and asphaltic concrete are approximately 1850 kg/m³ and 2300 kg/m³ respectively, an extra 'safety margin' of about 8% is included for each asphalt layer in order to ensure the presence of minimum (nominal) thicknesses throughout; the mean effective thicknesses are shown in brackets in the table. The bitumen content values in the 3rd column are also specified as prescribed nominal values.

200,000 m² and sometimes even more. The majority of projects thus had a length of 10-20 km with an average carriageway width of about 10 m. The overall average values for $\bar{x}$ ($\mu$) and the mean standard deviations $s$ ($\sigma$) shown in the second and third columns of Table 2 were calculated as follows. The random sample average ($\bar{x}$) and the (estimated) standard deviation ($s$) were determined for each project and each characteristic. Using the values for $\bar{x}$ and $s$, the mean value of the random sample averages ($\bar{\bar{x}}$) and the mean standard deviation ($\bar{s}$) were then calculated for each characteristic.
Where the construction work is performed to a proper standard and sufficient care is taken over quality control, the completed works meet the specified criteria and there are few, if any, penalties.

Table 3  Level of penalties with reference to number of controlled projects (1968-1975)

<table>
<thead>
<tr>
<th>Penalties not higher than guilders/t asphalt</th>
<th>Number of projects (%) (cumulative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>9</td>
</tr>
<tr>
<td>0.1</td>
<td>35</td>
</tr>
<tr>
<td>0.2</td>
<td>54</td>
</tr>
<tr>
<td>0.3</td>
<td>67</td>
</tr>
<tr>
<td>0.4</td>
<td>77</td>
</tr>
<tr>
<td>0.5</td>
<td>80</td>
</tr>
<tr>
<td>1.0</td>
<td>91</td>
</tr>
<tr>
<td>2.0</td>
<td>96</td>
</tr>
<tr>
<td>3.0</td>
<td>98</td>
</tr>
<tr>
<td>4.0</td>
<td>99</td>
</tr>
<tr>
<td>5.0</td>
<td>100</td>
</tr>
</tbody>
</table>

Current average cost per ton of asphalt in the Netherlands: approx. 50 guilders (approx. $25).
3 Statistical system

General

The system of quality control for asphalt pavements used up to now is, as we have seen, based largely on determination of the quality of individual cylindrical samples extracted by drilling from the completed pavement; the penalty is always related exclusively to the results of the tests carried out on these cylindrical samples. This essentially traditional method of quality control has certain drawbacks. Information which has become available in the past ten years or so (as in other technical sectors) shows that these drawbacks can be overcome by applying statistical methods.

In a statistical quality control system, interpretation of test results on the basis of mean values and standard deviations replaces the analysis of individual samples. The possibility of switching over to a fully statistical system was already discussed in the article by Van de Fliert and Brouwers (1968) referred to above; the principle referred to by them of relating the penalty level to the statistical excess percentage can also be considered as the basis of the new system which was described in detail in an earlier publication [8].

Testing system

Rational quality control must meet the following requirements:
a. The testing system must be such that acceptable works do not normally incur penalties while a penalty of any importance is only imposed on them in exceptional cases (producer's risk). On the other hand the testing system must be designed in such a way that 'bad' works do normally incur substantial penalties; this is very important, if only indirectly, as a preventive measure to encourage good quality control.
b. Because the quality of the works is determined on the basis of the results obtained by means of a non-selective random sample of limited size, 'chance' will have a relatively large influence on the results.

In others words the magnitude of the random sample average (x̄), and that of the standard deviation (s), are influenced by chance. The testing system must allow for such accidental deviations.

The term 'testing against variables' is used when the acceptance or rejection of a batch of products or the imposition of a penalty on the contract sum for a particular
project depend on variables such as the mean ($\bar{x}$) and standard deviation ($s$) of observations on a random sample taken from the batch or on core samples drilled from an asphalt pavement.

The aim is thus to keep the mean and standard deviation under control. A system of this kind has now been chosen for acceptance control of asphalt pavements and sand cement roadbases. In addition to the two parameters $\bar{x}$ and $s$, in testing these pavements and roadbases against measurable characteristics such as the voids content of asphaltic concrete and the compressive strength of sand cement, two further values must be known:
- a limit value or penalty limit $R$ below or above which a particular measurable characteristic must be defined as functionally ‘bad’;
- the uncertainty percentage which is still just acceptable; a value frequently encountered in industrial practice is 5%.

As mentioned above, the testing system must be designed in such a way that penalties may only occur in exceptional cases for acceptable works; a figure of 0.05 is often taken for this ‘producer’s risk’ – in other words a likelihood of approval (or, expressed differently, no imposed penalty) of 95%.

The following system is now used: the physical characteristic to be investigated is measured on a non-selective random sample of the prescribed size $n$. From the $n$ results obtained both the mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n},$$

and standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n} (\bar{x} - x_i)^2}{n - 1}}$$

are determined.

With the aid of these results, the testing parameter

$$\frac{R_{\text{max}} - \bar{x}}{s} \quad \text{or} \quad \frac{|R_{\text{min}} - \bar{x}|}{s}$$

is calculated, depending on whether the penalty limit $R$ is a maximum or minimum value. In one instance, namely determination of the bitumen content, where there is both a lower and an upper limit, both testing parameters must be determined.
If the testing parameter is greater than or equal to a constant $Q$ (quality index) no penalty is applied to the contract sum in respect of the characteristic concerned. If the criterion is not met, a penalty is imposed; its magnitude is dependent on the level of the result of the testing parameter; $Q$ is a constant whose magnitude is dependent on the size of the random sample and on the uncertainty percentage which is still just considered permissible, as well as on the probability of approval (or probability of no penalty) applicable for example to a batch of products which is still just acceptable (in practice often 95\%).

The $Q$ values will be calculated with the aid of the following simple formulae determined by Stange [2]:

$$Q = \frac{|\xi_0|}{1 + \varepsilon} \frac{\varepsilon}{|\xi_0|} \text{ with } \varepsilon = \frac{|\xi_{1-\alpha}|}{\sqrt{2n}},$$

in which $\xi_0$ = $\xi$-value applicable to the maximum permissible error percentage ($\delta$) or the maximum permissible percentage of 'bad' material.

$\xi_{1-\alpha} = \xi$ value applicable to the probability of approval or probability of no penalty for a batch with the maximum permissible percentage of 'bad' material; $\alpha$ is the threshold of unreliability: in practice a value of 0.05 is generally used in which case the probability of approval is 95\% ($n$ is the size of the random sample).

The percentages associated with the $\xi$ values will be found in tabular form in standard books on statistics (see [3], page 43).

If $\xi$ is negative, the table shows the fraction of observations smaller than $R$ and if $\xi$ is positive, the fraction of observations greater than $R$.

The formula for the testing parameters coincides well with the formula by which the percentage of observations in a normally distributed population which is smaller or larger than a given $R$ value, is determined, i.e.

$$\frac{R - \mu}{\sigma} = \xi$$

The difference now is that the mean $\mu$ and standard deviation $\sigma$ of the population are replaced by the mean $\bar{x}$ and standard deviation $s$ of the random sample.

If a number $n$ of cylindrical random samples are taken non-selectively from the same asphalt pavement, we shall naturally find, for the mean $\bar{x}$ and standard deviation $s$ of e.g. the asphaltic bitumen content, values which not only differ among themselves but also depart in a more or less random manner from the 'true' mean bitumen content $\mu$ and the 'true' standard deviation $\sigma$.

As we have already seen, account must be taken of the influence of these random
deviations on the criteria concerned; this is done by using Stange's formulae referred to above for determination of the $Q$ values. However, these criteria can in theory only be applied if the random samples are taken from populations with normal or practically normal distribution. A population with a distribution which departs significantly from that of a normally distributed population while still having the same mean ($\mu$) and standard deviation ($\sigma$) will have an uncertainty percentage which differs from the latter distribution. This may be a drawback for application of the criteria referred to above since the criteria concerned may become 'blurred' – with inevitable consequences for quality control. Research carried out by the American industry has, however, shown that there are seldom wide differences in uncertainty percentages between these distributions. Therefore in the overwhelming majority of cases the error made in applying these criteria to populations which deviate in varying degrees from the normal distribution will not be important [4]. Figure 2 shows as examples four histograms for different characteristics of layers of works chosen at random. It is apparent that the deviations from the normal distribution remain within reasonable limits.

![Histograms for certain characteristics of pavement layers chosen at random.](image)

Figure 2. Histograms for certain characteristics of pavement layers chosen at random.
By using the criteria defined above we have obtained a test procedure in which 'batches' (in this case pavement layers) which have identical uncertainty percentages but differ in terms of mean and standard deviation, have the same probability of approval (or of avoiding penalty imposition) (Fig. 3); this is very important in the road building sector where the asphalt mix production, among other factors, may differ widely from one contractor to another.

Figure 3. Two production processes which differ in terms of mean values and standard deviations but give the same percentage defectives.

**Size of random sample**

The size of the random samples chosen for statistical quality control of the various characteristics of asphalt pavements and sand cement roadbases is shown in Table 4 with reference to pavement area in m².

To ensure reasonably good differentiation between 'good' and 'bad' works, the random sample size per lot is set at 40 samples to determine the quality of the various characteristics of asphalt pavements and sand cement roadbases. Partly because determination of the asphaltic bitumen content of the different types of asphaltic concrete is extremely labour-intensive, and therefore expensive, a random sample size of 20 samples per lot is used for quality control of the asphaltic bitumen content; this number is necessary to obtain sufficient information on the quality of the completed work.
Table 4 Size of random samples for statistical quality control

<table>
<thead>
<tr>
<th>characteristic</th>
<th>size of random sample with ref. to pavement area in m²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20,000–200,000 m²</td>
</tr>
<tr>
<td>asphalitic bitumen content of bitumen bound gravel, open-textured asphaltic</td>
<td>20</td>
</tr>
<tr>
<td>concrete and dense asphaltic concrete</td>
<td></td>
</tr>
<tr>
<td>percentage voids of bitumen-bound gravel, open-textured asphaltic concrete and dense asphaltic concrete and dense asphaltic concrete</td>
<td>40</td>
</tr>
<tr>
<td>degree of compaction of sand asphalt</td>
<td>40</td>
</tr>
<tr>
<td>compressive strength of sand cement</td>
<td>40</td>
</tr>
<tr>
<td>thickness of the different layers</td>
<td>40</td>
</tr>
</tbody>
</table>

Road works with surface areas between 200,000 and 400,000 m² are divided into two, and works with a surface area greater than 400,000 m² into three sections (lots) of identical size. All the lots are the subject of individual quality control. To determine asphalitic bitumen content, 20 samples are used for each lot while 40 samples per lot are used to determine the quality of the other characteristics shown in Table 4.

For projects with a surface area of less than 20,000 m², statistical quality control using this system is considered too expensive. The obvious solution here is to assess the quality of such works by investigating individual samples taken for example for each 1000 m²; conclusions can then be drawn from the results of these tests of individual samples.

In the case of small works, however, a better alternative will normally be to ensure frequent and thorough control by the public authority of the performance of work by the contractor.

The new system always uses individual samples (taken non-selectively), i.e. each sample consists of a cylinder with a diameter of 0.10 m taken from the pavement perpendicular to the road surface; in the previous system on the other hand the samples consisted of two cylinders with a diameter of 0.10 m obtained by drilling from the pavement on transverse lines at intervals of 1.0 m, perpendicular to the road surface. In the new system of acceptance control it is desirable to use single samples, since the sampling method might otherwise introduce inaccuracies into the random sample sizes [5].

19
New criteria for quality control of asphalt pavements and sand cement roadbases

In defining the criteria for quality control of asphalt pavements and sand cement roadbases use was made, for the various characteristics, of the corresponding mean of the random sample averages (\(\bar{x}\)) and mean standard deviation (\(\bar{s}\)). These values have been calculated with reference to the quality control results for several hundred projects in the years 1968-1978 (see Table 2, columns 3 and 4).

In these calculations, the values for \(\bar{x}\) and \(\bar{s}\) only required modification in one single instance on the basis of practical data and information on desired and feasible quality levels. It was assumed that the requirements for the level (\(\mu\)) and extent of dispersion (\(\sigma\)) (i.e. mean values calculated from the results not only of good but also of moderate and even poor works) must still be just acceptable to the directorate and considered reasonably feasible by the contractor ('standard job'). This seems a better assumption than to take values for \(\mu\) and \(\sigma\) calculated solely from the results of moderate and good works.

In determining the criteria on the basis explained above, the \(Q\) values were always defined in such a way that where the mean level and standard deviation for the various characteristics just met the required standards, there would be a 95% probability of no penalty in respect of each characteristic. This principle can be summarized as follows: in the newly developed system of acceptance control, the level and dispersion values used are taken from data for a large number of projects with the criteria fixed in such a way that when these values are respected the likelihood of a penalty is slight.

Composition and density of asphalt mixes and strength of sand cement

Columns 3 to 9 of Table 2 show the following data for the different characteristics:
- the values determined in the manner outlined above for the mean (\(\mu\)) and standard deviation (\(\sigma\)) used – with reference to the results achieved in practice or the values prescribed for \(\mu\) – as the basis for defining the criteria;
- the penalty or quality limits (\(R\)) now fixed on the basis of practical results and statistical considerations and representing, by definition, the boundary between 'good' or 'satisfactory' and 'bad' or 'not satisfactory' work:
- the \(\xi\) values calculated as \(\frac{R - \mu}{\sigma}\) and the maximum permissible error percentages (\(\delta\)) determined therefrom with the aid of tables, using the method of calculation described above (see [3], page 43):
- the subsequently determined \(Q\) value which shows that with the given random sample size (40 or 20) there is a 95% acceptance probability for a characteristic with the maximum permissible error percentage;
- the definitive \(Q\) value determined by rounding-off or fixed for reasons of uniformity.
The following explanation is appropriate here.

The $Q$ value determined by rounding-off or fixed for reasons of uniformity functions in the general criterion as follows:

No penalty where

$$\frac{|R_{\text{min or max}} - \bar{x}|}{s} \geq Q.$$  

Penalty imposed where

$$\frac{|R_{\text{min or max}} - \bar{x}|}{s} < Q$$

In the latter instance the excess percentage associated with the calculated value for the penalty figure

$$\frac{|R_{\text{min or max}} - \bar{x}|}{s}$$

will be found in the table for the normal distribution [3]. For quality control of the asphaltic bitumen content of the different types of asphaltic concrete, a double criterion will be applied in the following manner:

No penalty where:

$$\frac{|R_{\text{min}} - \bar{x}|}{s} \geq Q \quad \text{and} \quad \frac{R_{\text{max}} - \bar{x}}{s} \geq Q$$

A penalty is on the other hand imposed where:

$$\frac{|R_{\text{min}} - \bar{x}|}{s} < Q \quad \text{and/or} \quad \frac{R_{\text{max}} - \bar{x}}{s} < Q$$

If in both cases the value of the test parameter is lower than $Q$, the corresponding excess percentages $B_1$ and $B_2$ must be added together.

**Layer thickness**

As regards the layer thickness, the penalty system is only applied for three separate thicknesses, i.e.

- the overall thickness of the different layers of asphaltic concrete, i.e. of the bitumen-
bound gravel, open-textured asphaltic concrete and frequently also dense asphaltic concrete which are of primary importance to the entire dimensional characteristics of the pavement;
- the thickness of the roadbase where it does not consist of bitumen-bound gravel (i.e. sand cement or sand asphalt) which plays a separate role in the dimensions of the pavement;
- the thickness of the temporary or permanent surface layer of the pavement with the emphasis placed on durability of this layer from the angle of direct mechanical stress and maximum particle size.

Apart from the pavement surface layer, the total thickness of the various layers of asphaltic concrete is also determined; it is therefore naturally possible to compensate at any time on the site a shortfall caused for one reason or another in the thickness of a particular layer, by applying additional thickness to one or more of the upper layers.

Statistical interpretation is effected wherever possible, i.e. in so far as the number of samples taken from the pavement or the results obtained by local measurement allow this. In cases (e.g. service roads, local roads, cycle tracks and access roads) where the number of cylindrical samples to be taken from the pavement is insufficient for statistical evaluation, the present system of penalties with testing of individual samples, is maintained in respect of layer thickness (see Table 1). The present system of penalties can also be applied to reconstruction projects where the nature of the work is such (wide variations in layer thickness) that statistical interpretation would be unrealistic.

To ensure convenient and uniform test criteria for the overall layer thickness of asphaltic concrete structures, the thickness of sand asphalt or sand cement roadbases and the thickness of temporary or permanent surface layers, the same criterion is applied in all possible cases (see Table 2) \( Q = 1.40 \).

To ensure that, with the set \( Q \) value of 1.40, works with correct layer thickness \( (\mu) \) (i.e. the layer thickness for which the contractor is to be paid), and a correct standard deviation \( \sigma \), are 95% sure to avoid penalty with a random sample size of 40 core samples, it is possible, using the known basic data, to calculate the penalty limit \( R_{\text{min}} \) in the various cases from the following formulae \( (R_{\text{min}} \) being a completely arbitrary number):

\[
1.40 = \frac{\left| \bar{z}_d \right|}{1 + \varepsilon} - \frac{\varepsilon}{\bar{z}_d} \quad \text{and} \quad \varepsilon = \frac{1.85}{\sqrt{80}} \quad \text{or:}
\]

\[
\bar{z}_d = 1.78 \quad \text{(permissible uncertainty percentage: 3.8%)}
\]

\[
\frac{R_{\text{min}} - \mu}{\sigma} = -\bar{z}_d \quad \text{or} \quad R_{\text{min}} = \mu - 1.78 \sigma
\]
Table 2 shows penalty limits ($R_{\text{min}}$) calculated with the aid of these formulae for the overall thickness of the various layers of asphaltic concrete and for the thickness of the separate layers of dense asphaltic concrete, open-textured asphaltic concrete, sand asphalt and sand cement.

**Atypical values**

In incidental cases a value may occur (in a random sample of 20 or 40 results) which does not really seem to belong to the population. These values may for example fall outside the ‘3σ limit’. Such results which are attributable to a variety of causes and are not always due to careless work by the contractor, have a relatively large influence on the standard deviation ($s$) and thus also on the value of the test parameter

$$\frac{|R_{\text{min}} - \bar{x}|}{s} \quad \text{or} \quad \frac{R_{\text{max}} - \bar{x}}{s}.$$  

To give the contractor the benefit of the doubt, the following rule is applied:

If one (and not more than one) of the 20 measured results for the asphaltic bitumen content of the various types of asphaltic concrete or one of the 40 results for the voids content of the various types of asphaltic concrete, the degree of compaction of the sand asphalt or the compressive strength of the sand cement, fall outside the limit values shown below, this result will be treated as a value which does not belong to the population (and is therefore atypical: see Table 5).

**Table 5 Limit values for atypical results**

<table>
<thead>
<tr>
<th>characteristic</th>
<th>limit value for atypical results</th>
</tr>
</thead>
<tbody>
<tr>
<td>asphaltic bitumen content of bitumen-bound gravel, open-textured asphaltic concrete and dense asphaltic concrete</td>
<td>$V_b \pm 1.0%$*</td>
</tr>
<tr>
<td>Voids of:</td>
<td></td>
</tr>
<tr>
<td>bitumen-bound gravel</td>
<td>12%</td>
</tr>
<tr>
<td>open-textured asphaltic concrete</td>
<td>11%</td>
</tr>
<tr>
<td>dense asphaltic concrete</td>
<td>10%</td>
</tr>
<tr>
<td>degree of compaction of sand asphalt</td>
<td>92%</td>
</tr>
<tr>
<td>compressive strength of sand cement</td>
<td>1.0 MN/m$^2$</td>
</tr>
</tbody>
</table>

Application of the ‘3σ rule’ leads, with some rounding-off, to these limit values; because of the relatively wide dispersion, the ‘3σ rule’ is not applicable to the compressive strength of sand cement; an arbitrary limit value of 1.0 MN/m$^2$ is therefore taken, below which the test results can scarcely be considered reliable.

* $V_b = \text{prescribed value}$
Allowance for atypical values in the case of layer thickness control appears less appropriate for a variety of reasons, especially as the accuracy of a value which appears to be seriously inadequate can normally be verified by direct and objective means.

**Test characteristics**

Figures 4 to 6 inclusive show the operating characteristic curves in respect of the criteria defined above for assessing the bitumen content and voids percentage of the various kinds of asphaltic concrete, the degree of compaction of sand asphalt and the compressive strength of sand cement.

Details of the method for calculating these characteristic curves will be found in handbooks. The test characteristic shown in Figure 5 is also applicable to the acceptance criteria for total thickness of the different layers of asphaltic concrete, and for the layer thicknesses of dense and open-textured asphaltic concrete, sand asphalt and sand cement. These curves show the relationship between the probability of approval ($P_a$) or likelihood of no penalty, and the percentage of 'bad' material or the error percentage attributable to a layer of asphaltic concrete, sand asphalt or sand cement. As regards the characteristic curve for the bitumen content of various types of asphaltic concrete (see Figure 4) it should be noted that precise determination is not possi-

![Figure 4](image-url)
ble, partly because of the fact that a dual criterion is applied for the determination of the asphaltic bitumen content. In particular for relatively low error percentages, between 0.5 and 1.5%, the probability of no penalty being imposed will be somewhat lower than the characteristic curve suggests.

The different figures clearly show that penalties will only be incurred in exceptional cases by good work while poor work will almost always be penalized. For example, if the error percentage in respect of the voids content of a layer of bitumen-bound gravel is 1%, the likelihood of no penalty being imposed is 99.9%; if on the other hand the error percentage is 8% (for example for work with a mean value of 6.7% and a standard deviation of 2.0%), the likelihood of penalties being justifiably imposed is very high, i.e. in the order of 80% (see Figure 5).

As we have already seen, a penalty may exceptionally be imposed on the contract sum where the work is in fact good; this is known as the producer's risk. On the other hand there is also a consumer's risk: it may happen, although again this is exceptional, that no penalty is imposed on work which is definitely bad. For example, if the error percentage in respect of the voids content of a layer of bitumen bound gravel is 10% (e.g. for work with a mean value of 6.9% and a standard deviation of 2.0%), the likelihood of no penalty is 9%; in other words 1 out of 11 such works will escape any penalty. It should of course be noted that in the statistical method of quality control of a road project, individual characteristics are not assessed separately; on the contrary a number (generally about 10) of characteristic parameters (bitumen content and density of asphalt, strength of sand cement, layer thicknesses of 3 or 4 different pavement layers) are determined.

This in turn means that where a penalty is wrongly imposed on good work or alternatively no penalty incurred by bad work, the error will only apply to one part of the work and never to the work as a whole so that the consequences which may arise in exceptional cases affect only that particular part of the work and never the project in its entirety.

In determining the various criteria for the purpose of acceptance control, the analysis is in all cases based on values for the level and extent of dispersion which the contractor can reasonably be expected to meet.

It is open to question whether, in the event of reduction of the minimum standards placed on the different characteristics, the work may still just be acceptable. However, a reduction of these minimum requirements necessarily entails a higher error percentage which increases the consumer's risk; in that event the likelihood of poor work escaping any penalty is higher. From the standpoint of good quality control, a reduction of the set minimum standards must therefore be treated as impermissible. Calculations have also been made to determine the approximate percentage probability of relatively low or high excess percentages in respect of certain mean values ($\mu$) and standard deviations ($\sigma$) for the various characteristics.

For details of the calculation of these probabilities, reference should be made to standard statistical handbooks. The results of the calculations show that if, in excep-
tional cases, a penalty is charged on good work that penalty will be relatively low, i.e.
less than 2 to $2\frac{1}{2}\%$ of the total cost of the layer, since the excess percentage will prac-
tically never be higher than the 10\% allowed for the asphaltic bitumen and voids
content for the various kinds of asphaltic concrete and the 15\% for the degree of
compaction of sand asphalt and compressive strength of sand cement. Again in the
case of works which are slightly below the minimum requirements for level and dis-
persion, the likelihood of an excess percentage (depending on the characteristic) of
more than 10\% or 15\% is extremely low.
A high probability of an excess percentage of more than 25\% in fact only occurs
– and then quite rightly – in the case of bad work with error percentages of 20\% or
more.
Considerable likelihood of excess percentages between 15 and 25\% for asphaltic
bitumen content and voids content occurs in the case of works which fall distinctly
short of the set minimum standards relating to mean and standard deviation. In gen-
eral these are works with error percentages between 12 and 20\%.

Penalty criteria

When these criteria are applied it is reasonable to assume that this statistical system
is less stringent than the system applied hitherto, especially for percentages which are
only marginally in excess of the penalty limits. It has been found in practice that
penalties on many works fall precisely within this area.
Critical consideration of the relationship between shortfall on quality and the result-
ing, desirable compensation charges or penalties, leads us to conclude that penalties
under the present system are in fact relatively low.

Penalties

In the light primarily of experience of the non-statistical quality control system used
during the past ten years, the relationship between the scale of the penalties $K$ and
the percentage of 'unsatisfactory' or 'poor' test results $B$ in the statistical system has
been determined as follows:

$K = 0.3 \times B - 1.0$ for bitumen content and voids content of asphaltic concrete, and
$K = 0.3 \times B - 2.0$ for layer thicknesses, degree of compaction of sand asphalt and
compressive strength of sand cement.

$K$ is expressed here as a percentage of the true value of the layer concerned, while $B$
is the percentage of the overall work which, on the basis of the calculation (from the
mean value $\bar{x}$ and standard deviation $s$), can be considered to fall below the prescribed
quality (penalty) limit $R$ in respect of a particular characteristic ($B = \% \text{ defectives}^\prime$).
Figure 5. Characteristic curve for determination of voids content of asphaltic concrete.

Figure 6. Characteristic curve for determination of degree of compaction of sand asphalt and compressive strength of sand cement.
**Introduction of the new statistical system**

In discussions with the committee of experts of the road building contractors organizations, a wish was expressed for more information on the consequences of the new system before it was actually introduced. It was agreed that as an initial step a number of projects (10 to 20) would be assessed by both systems, i.e. the traditional non-statistical and the new statistical system. The results would then be compared and discussed with a view to possible adjustment of the statistical system.

**Costs and benefits**

It must be remembered that neither quality control during implementation nor acceptance control can in themselves give complete certainty as to the overall average quality of a particular project or prevent instances of poor quality. Investigations are and remain based on arbitrary, random sample checks. It also practically impossible to obtain optimum quality at the lowest possible cost to the authority which has commissioned the works. The aim of the penalty provisions cannot therefore be primarily to provide equitable compensation for lower quality. All rules laid down on a largely theoretical basis with that aim in mind would be influenced by so many factors (subsoil, traffic load, maintenance methods, weather conditions etc.) that a precise approximation is impossible. The introduction of an apparently watertight system based on cost-benefit analysis, would therefore be inequitable to the contractors who perform the work. It is, however, particularly important to have some knowledge of the costs and benefits arising in this context. We refer to the costs resulting from a shorter useful life due to poor quality, and to the benefits flowing from the proceeds of the penalties imposed because of quality shortcomings.

A special study of this problem was made by Brouwers in 1974; it was hitherto only available in Dutch [10], but now appears as chapter 4 of this publication.

**Control of surface characteristics**

An important area of acceptance control relates to the surface characteristics of evenness and skid-resistance. Measurements of evenness and skid resistance are made by methods which are not entirely statistical. The fact that it is not easy to express traffic safety in statistical figures plays a part here. In the first instance only 30% of the road length is checked for evenness and skid-resistance in order to limit the number of measurements. The location of the measuring points is fixed at random. How-
ever, if the results of this first series of measurements, and the appearance of the surface, give reason to suppose that there are large-scale shortcomings on quality, especially as regards skid-resistance, further measurements are carried out; should this be necessary, the entire surface of the completed work is controlled.

**Evenness**

Until 1975 the evenness was always checked by using a normal or rolling straightedge with a length of 3 m. When deviations from the even profile of more than 3 mm were found in a measured section with a length of 100 m, penalties were imposed. If deviations of more than 5 mm were found more than once, the evenness had to be corrected by shaving – naturally at the contractor’s cost. Since 1976 the viagraph has been used instead of the rolling straightedge to check the evenness of the carriageway surface. A penalty is imposed if the deviation percentage $C_s$ is greater than 2. Deviations of more than 5 mm, measured with the viagraph, necessitate correction by shaving at the responsibility of the contractor concerned. In special cases where the irregularity is such that correction by shaving will not result in a completely even surface, the contractor is required by special contract provisions to adjust the surface by laying an extra surface course of 40 mm on the completed road surface.

![Figure 7. Viagraph for measuring evenness of road surfaces.](image)
Skid-resistance

Since as long ago as 1967 it has been stipulated in the Netherlands that before trunk roads are opened to traffic, new pavements must be checked for skid-resistance. This is done with a standard measuring vehicle with a retarded wheel (86% slip), a wet surface and a measuring speed of 50 km/h. Penalties are imposed if the measured coefficient of friction is lower than 0.52.* However, if the value is less than 0.45* the contractor must also correct the surface of the asphaltic concrete by treatment with white spirit and crushing sand. This treatment adequately removes surplus bitumen from the surface until a coefficient of friction of at least 0.52 is obtained. However, it is very expensive and the provision in the contract thus has a good deterrent effect. New roads with smooth surfaces are thus rarely encountered today and the completed surface normally easily meets the requirement regarding skid-resistance. Other methods of improvement have also been applied recently, i.e. sand-blasting or milling.

Figure 8. Standard vehicle for measuring skid resistance of road surfaces.

* These limit values are specified since January 1978. With respect to international standardization of skid-resistance test tyres as from 1978 the measuring vehicles of the SRL have been provided with standard test tyres of a type different from the one used since 1958. Starting from the original criteria the numerical values for the specifications on skid-resistance of road surfaces have been changed on the basis of comparative measurements with both types of tyres. Before 1978 the specified limits were 0.56 and 0.51 instead of 0.52 and 0.45 respectively.
One stipulation is particularly important in this connection. For more than ten years Dutch road contract specifications have stated that warm asphaltic concrete surface layers must be spread during rolling with approx. 2 kg fine stone chippings (2-5 mm) per m$^2$. This treatment effectively prevents initial smoothness of new asphaltic concrete surfaces due to excess bitumen content in the surface.
4 Relationship between the penalty system and the necessary compensation costs

In this final chapter calculations are given for the layer thickness, voids and asphaltic bitumen content of an asphalt pavement; these calculations show that the penalties are substantially lower than the compensation costs. In the case of the layer thickness reference is made to a pavement structure which is frequently used in practice and the consequences of divergent layer thickness distributions (both too low average thickness and excessive standard deviations) on the service life of the pavement are calculated. Comparisons are drawn with the average service life of the standard pavement structure determined by a design method.

Similar calculations are effected for the voids and bitumen content; the reduction in service life can again be calculated with reference to an asphalt structure with voids equivalent to the mean voids content of all works to which penalties have been applied, or with a bitumen content equivalent to the desired bitumen content.

The following formulae are used for calculation of the penalties:

\[ K = 0.2B \] for the layer thickness;
\[ K = 0.3B \] for asphaltic bitumen content and voids.

When this study was carried out, about one year before development of the statistical system described above was fully completed, it was not yet known which specific formulae would be proposed in the final version. Although the latter do in fact differ somewhat from the formulae reported earlier (i.e. \( K = 0.3B - 2.0 \) and \( K = 0.3B - 1.0 \)) the original formulae have been maintained in this chapter, as Tables 3 to 7 would otherwise have had to be revised. The differences concerned are only slight so that this has no great influence on the conclusions drawn.

Layer thickness

The damage resulting from insufficient layer thickness can be quantified on the basis of design data establishing a relationship between the layer thickness and traffic parameters. Use is made below of the 'design formula of the AASHO test' which establishes the relationship between the number of load repetitions \( n \) up to the attainment of a given serviceability index \( p \), and the thickness, expressed as the thickness index \( D \) for a given wheel load \( P \) (tons). This formula, valid for \( p = 1.5 \) - a low level
of, in particular, longitudinal evenness of the carriageway at which reconstruction becomes necessary, is as follows:

\[ n = \frac{10^{5.93}(D + 1)^{9.36}}{(4.41P + 1)^{4.79}} \]

In this formula \( D = \sum c_i h_i = c_1 h_1 + c_2 h_2 + \ldots \), an equivalent pavement thickness formed by summation of the layer thicknesses of the different pavement layers (\( h \) in cm), each multiplied by the value coefficient \( c \).

For the reference wheel load \( P = 5 \) tons, this formula becomes

\[ n = 0.2528(D + 1)^{9.36} \]

or \( D = 1.1583n^{0.107} - 1 \), where \( n \) is the equivalent number of load repetitions of the 10 ton axle load (or equivalent 5 ton wheel load). This design formula naturally only applies under conditions similar to those for the AASHO test as regards the underlying subsoil and the environment.

For a more detailed description of the AASHO test reference should be made to specialized publications, e.g. the paper by Van de Fliert and Brouwers [9].

The example chosen is the standard structure of a primary road (non-motorway) based on:

\[
\begin{align*}
4 \text{ cm coarse dense asphaltic concrete} & \quad \text{with } c = 0.173 \\
4 \text{ cm open-textured asphaltic concrete} & \quad \text{with } c = 0.12 \\
18 \text{ cm (3 x 6 cm) bitumen-bound gravel} & \quad \text{with } c = 0.02 \\
\end{align*}
\]

There are thus 26 cm asphaltic concrete in all (nominal layer thickness) applied to a compacted subgrade with a thickness of 50 cm and \( c = 0.02 \).

The thickness index of this structure is

\[ D = 8 \times 0.173 + 18 \times 0.12 + 50 \times 0.02 = 4.55 \]

On the basis of the invoicing clauses in contract specifications with penalty provisions, the effective mean layer thickness of the bituminous structure is as follows with a mean unit density (normal value) of 2.33:

\[
\begin{align*}
\frac{2.50}{2.33} \cdot 4 &= 4.3 \text{ cm dense asphaltic concrete} \\
\frac{2.50}{2.33} \cdot 4 &= 4.3 \text{ cm open-textured asphaltic concrete} \\
\frac{2.50}{2.33} \cdot 18 &= 19.3 \text{ cm bitumen-bound gravel} \\
\end{align*}
\]

Total = 28 cm asphaltic concrete.
The thickness index of the average structure is therefore:

\[ D = 8.6 \times 0.173 + 19.3 \times 0.12 + 50 \times 0.02 = 4.80 \]

On this pavement with \( D = 4.80 \), the dimension formula shows that \( n = 3.536 \times 10^6 \) (equivalent 10 ton axle) load repetitions are permissible before the serviceability index \( p = 1.5 \) is reached.

The service life of this pavement in years can be calculated on the basis of two further factors: the number of equivalent 10 ton load repetitions in the year of construction and the annual percentage growth in this number. We have taken for these factors \( n_o = 84,000 \) and \( \alpha = 7\% \) respectively, i.e. an equivalent annual total and a mean growth percentage which have recently been used as the characteristic figures for primary roads.

The service life in years can now be calculated (for example) graphically with the aid of a graph showing the cumulative total of axle loads as a function of the service life in years for the applicable (continuous) growth percentage. It is thus calculated that the structure under consideration with \( \Sigma n_{eqs_t} = 3.536 \times 10^6 \) has a service life \( T \) of 20.3 years. On the basis of the relationship \( D = \Sigma \frac{c_i}{h_i} = f(n)_{p,p} \) and \( T = f_{n_o,a} \), i.e. \( D = f(T) \), the service life in years can now be calculated for every pavement thickness.

The reduction in layer thickness proportional to the design thicknesses is distributed over the asphalt layers so that with the asphalt thickness \( H \) in cm:

\[ D = (4.80 - 1.00) \frac{H}{28} + 1.00. \]

On the basis of a normal population distribution, the layer thickness of the pavement is entirely determined by the mean layer thickness \( \mu \) and standard deviation \( \sigma \). 28 cm is taken as the feasible average for this ‘normal’ structure; this is the layer thickness which, in the closest possible approximation to the calculated daily quantity of 25 kg/m² per cm layer thickness, can be expected for asphalt with a relatively high unit density. 2.0 cm is taken as the feasible standard deviation for the total layer thickness. Both assumptions can be considered realistic on the basis of the results obtained. It is now also easy to show, on the basis of the normal distribution, the dispersion in layer thickness over this total ‘standard’ pavement and hence to express the dispersion in terms of service life.

The calculations have been compiled in Table 6.

The significance of the ‘nominal layer thickness’ is now clearly apparent: 84.1% of the pavement surface has a thickness equal to or greater than the nominal thickness of 26 cm, and therefore a service life equal to or greater than 16 years.

It is also more in accordance with road-building practice to take as the effective service life of the pavement the life at the nominal layer thickness rather than at the mean layer thickness. Premature maintenance of the pavement will certainly be necessary.
### Table 6  Dispersion in layer thickness of the ‘standard’ pavement and service life of parts of the pavement

<table>
<thead>
<tr>
<th>asphalt pavement layer thickness $H$, cm</th>
<th>pavement surface, % of total $(Q)$</th>
<th>service life of pavement</th>
<th>asphalt layer thickness, cm</th>
<th>equivalent no. 10 t axles</th>
<th>years, $T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&gt; 28$</td>
<td>50</td>
<td>$&gt; 28$</td>
<td>$&gt; 4.80$</td>
<td>$&gt; 3,536 \cdot 10^6$</td>
<td>$&gt; 20.3$</td>
</tr>
<tr>
<td>$27 - 28$</td>
<td>19.1</td>
<td>$28$</td>
<td>$4.80$</td>
<td>$3,536 \cdot 10^6$</td>
<td>20.3 (20)</td>
</tr>
<tr>
<td>$26 - 27$</td>
<td>15.0</td>
<td>$27$</td>
<td>$4.664$</td>
<td>$2,833 \cdot 10^6$</td>
<td>18</td>
</tr>
<tr>
<td>$25 - 26$</td>
<td>9.2</td>
<td>$26$</td>
<td>$4.529$</td>
<td>$2,258 \cdot 10^6$</td>
<td>15.7 (16)</td>
</tr>
<tr>
<td>$24 - 25$</td>
<td>4.4</td>
<td>$25$</td>
<td>$4.393$</td>
<td>$1,789 \cdot 10^6$</td>
<td>13.5</td>
</tr>
<tr>
<td>$23 - 24$</td>
<td>1.7</td>
<td>$24$</td>
<td>$4.257$</td>
<td>$1,409 \cdot 10^6$</td>
<td>11.5</td>
</tr>
<tr>
<td>$22 - 23$</td>
<td>0.5</td>
<td>$23$</td>
<td>$4.121$</td>
<td>$1,103 \cdot 10^6$</td>
<td>9.6 (9.5)</td>
</tr>
<tr>
<td>$&lt; 22$</td>
<td>0.1</td>
<td>$22$</td>
<td>$3.986$</td>
<td>$8,580 \cdot 10^5$</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$21$</td>
<td>$3.850$</td>
<td>$6,627 \cdot 10^5$</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$20$</td>
<td>$3.714$</td>
<td>$5,081 \cdot 10^5$</td>
<td>5.2 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$19$</td>
<td></td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

### Table 7  Layer thickness dispersion of 5 pavements with excessively low mean thickness or high standard deviation and effective life of these pavements

<table>
<thead>
<tr>
<th>variant</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean layer thickness $\mu$, cm</td>
<td>27</td>
<td>26</td>
<td>24</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>standard deviation $\sigma$, cm</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>% of pavement surface with layer thickness:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$&gt; 28$ cm</td>
<td>30.9</td>
<td>15.9</td>
<td>50.0</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>$27 - 28$ cm ($= 28$)</td>
<td>19.1</td>
<td>15.0</td>
<td>12.9</td>
<td>9.9</td>
<td>84.1</td>
</tr>
<tr>
<td>$26 - 27$ cm ($= 27$)</td>
<td>19.1</td>
<td>19.1</td>
<td>15.9</td>
<td>11.9</td>
<td>84.1</td>
</tr>
<tr>
<td>$25 - 26$ cm ($= 26$)</td>
<td>15.0</td>
<td>19.1</td>
<td>15.0</td>
<td>9.3</td>
<td>8.2</td>
</tr>
<tr>
<td>$24 - 25$ cm ($= 25$)</td>
<td>9.2</td>
<td>15.0</td>
<td>19.1</td>
<td>6.7</td>
<td>6.8</td>
</tr>
<tr>
<td>$23 - 24$ cm ($= 24$)</td>
<td>4.4</td>
<td>9.2</td>
<td>19.1</td>
<td>4.4</td>
<td>5.3</td>
</tr>
<tr>
<td>$22 - 23$ cm ($= 23$)</td>
<td>1.7</td>
<td>4.4</td>
<td>15.0</td>
<td>2.5</td>
<td>3.9</td>
</tr>
<tr>
<td>$21 - 22$ cm ($= 22$)</td>
<td>0.5</td>
<td>1.7</td>
<td>9.2</td>
<td>1.3</td>
<td>2.6</td>
</tr>
<tr>
<td>$20 - 21$ cm ($= 21$)</td>
<td>0.1</td>
<td>0.5</td>
<td>4.4</td>
<td>0.6</td>
<td>1.8</td>
</tr>
<tr>
<td>$19 - 20$ cm ($= 20$)</td>
<td>0.1</td>
<td>1.7</td>
<td>0.6</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>$&lt; 19$ cm ($= 19$)</td>
<td>0.1</td>
<td>0.6</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>effective layer thickness, cm</td>
<td>25</td>
<td>24</td>
<td>22</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>effective life $T$, years</td>
<td>13.5</td>
<td>11.5</td>
<td>8</td>
<td>13.5</td>
<td>11.5</td>
</tr>
</tbody>
</table>
for that part of the pavement which is thinner than 25 cm, i.e. 6.7% of the total pavement area.

In the following analysis, it is always assumed that the greatest layer thickness of the thickness ranges shown in the table is the determining factor as far as service life is concerned. It is now possible to determine this life distribution for pavements with other layer thickness distributions, i.e. different mean layer thickness values \( \mu \) and standard deviations \( \sigma \), and, on the basis of a comparison with the standard structure, to determine the cost incurred for premature reconstruction and maintenance.

For this purpose the same distribution over the layer thickness groups shown in the table must be calculated, the service life of 8.41% of the pavement surface being taken as the effective service life.

The calculations (see table 7) were made for five variants with mean layer thickness deviation \( \mu \) and standard deviation \( \sigma \).

For the decisive part of the six construction variants taken from tables 6 and 7, Figure 9 shows the (standard) distributions of the pavement surface as a function of the layer thickness.

![Figure 9. Distribution of road surface as a function of layer thickness and service life.](image)
thickness: the service life calculated from the layer thickness is also shown.
The following assumptions are made for the purpose of calculating the cost of pre­mature reconstruction and maintenance:
At the end of the effective life \((T\text{ years})\), an asphalt layer must be laid down; the cost of this work is set at \(f A\) per m\(^2\) road surface.
In the normal structure, this asphalt layer will serve for an (extended) service life of 10 years (i.e. \(16 + 10 = 26\) years after laying the pavement).
Since the effective life of the structure is only \(T'\) years instead of \(T\) years, this asphalt layer must be applied \((T - T')\) years sooner. This asphalt layer must have a greater thickness than under normal circumstances so that the service life is extended to \(10 + (T - T')\) years.
At that point in time the original situation is restored. The updated cost per m\(^2\) is therefore:

\[
\frac{(1 + j)^{T-T'} - 1}{(1 + j)^T} \cdot A + \frac{B}{(1 + j)^T}
\]

\(j\) = discount rate equated with the interest rate on long-term public notes; a figure of 0.08 (8%) is used.

\(A\) = cost of a 4 cm asphalt layer; the total cost (inclusive of ancillary charges) is set at \(f8/\text{m}^2\) road surface.

\(T\) = effective life of the extended structure, i.e. 16 years in this example.
The first term is the amount that must now be earmarked to apply the asphalt layer \((T - T')\) years sooner than normal.

\(B\) = cost of the extra thickness required for the asphalt layer per m\(^2\) to ensure that it has a life of \((T - T')\) years longer than 10 years.
The cost of the extra layer thickness of the asphalt layer to be applied after \(T'\) years to ensure a service life of \(10 + (T - T')\) years is calculated on the basis of a linear relationship between service life and layer thickness, so that:

\[
B = \frac{16 - T'}{10} \cdot 4\text{ cm}
\]

(rounded off to whole cm).
Although this linearity - on the basis of design relationships (in effect reconstruction) - in fact results in excessive additional thickness, it must be remembered that the application of the asphalt layer also takes place by reason of factors which are not related to design and tend rather to be proportional to \(n\) (the 'improvement' is a hybrid of maintenance and actual reconstruction for design reasons).
The second term is thus the amount that must now be earmarked for the additional layer thickness to be applied after \(T'\) years. Finally the costs are expressed as a per-
centage of the total pavement costs set at f 20.-, so that the following improvement costs to be compensated are arrived at:

\[
\left\{ \frac{(1.08)^{16} - T\cdot 1}{1.08^{16}} : \frac{16 - T'}{10} \cdot \frac{4}{20} \right\}\% \\
\]

In addition to these costs necessitated by premature improvement, the cost of the associated increase in maintenance can also be assessed, again in comparison with a normal structure. For this purpose, the difference in pavement area (as % of the total) must be calculated on the basis of the tables for the various thickness and service life groups. Additional costs will therefore be incurred for this purpose at the end of the effective life. Assuming that these costs are identical to the cost of normal reconstruction, i.e. f 8 per m\(^2\) – which is plausible given that the non-selective distribution of the damage often entails the improvement of a substantially greater road surface area than is strictly necessary – the following percentage of the total pavement costs is arrived at:

\[
\frac{8}{20} \sum \frac{Q - Q_0}{1.08^T} \text{ percent} \\
\]

\(Q\) is the percentage of the surface area in the layer thickness group for which maintenance is necessary prior to reconstruction. \(Q_0\) is the percentage in the same layer thickness group for a normal structure.

The calculations for both amounts in respect of the five variants are shown below:

Structure I \((\mu = 27 \text{ cm}, \sigma = 2 \text{ cm})\)

improvement: \(\left\{ \frac{(1.08^{2.5} - 1)8}{1.08^{16}} + \frac{1}{1.08^{13.5}} \right\}\frac{100}{20} = 4.1\%\)

maintenance: \(\frac{8}{20} \left\{ \frac{4.4 - 1.7}{1.08^{11.5}} + \frac{1.7 - 0.5}{1.08^{9.5}} + \frac{0.6 - 0.1}{1.08^8} \right\} = 0.8\%\)

total 4.9\% of pavement costs

Structure II \((\mu = 26 \text{ cm}, \sigma = 2 \text{ cm})\)

improvement: \(\left\{ \frac{(1.08^{4.5} - 1)8}{1.08^{16}} + \frac{2}{1.08^{11.5}} \right\}\frac{100}{20} = 9.1\%\)
maintenance: \[
\frac{8}{20} \left\{ \frac{4.4 - 0.5}{1.08^{8.5}} + \frac{1.7 - 0.1}{1.08^8} + \frac{0.6}{1.08^6.5} \right\} = 1.2\% 
\]

total 10.3% of pavement costs

Structure III \((\mu = 24 \text{ cm}, \sigma = 2 \text{ cm})\)

improvement: \[
\left\{ \frac{(1.08^8 - 1)8}{1.08^{16}} + \frac{3}{1.08^8} \right\} \cdot \frac{100}{20} = 18.1\% 
\]

maintenance: \[
\frac{8}{20} \left\{ \frac{4.4}{1.08^{6.5}} + \frac{1.7}{1.08^5} + \frac{0.6}{1.08^4} \right\} = 1.75\% 
\]

total 19.9% of pavement costs

\textit{N.B.} As an alternative form of improvement, the application of a new 4 cm surface layer can be calculated; the cost is then:

\[
\frac{100}{1.08^8} \cdot \frac{8}{20} = 24\% 
\]

It must be remembered that this is an additional operation which should not have been necessary and in the case of which allowance must also be made for compensation in respect of prejudice to the road user and road authority.

Structure IV \((\mu = 28 \text{ cm}, \sigma = 3 \text{ cm})\)

improvement: as structure I, 4.1%

maintenance: \[
\frac{8}{20} \left\{ \frac{4.4 - 1.7}{1.08^{11.5}} + \frac{2.5 - 0.5}{1.08^{9.5}} + \frac{1.3 - 0.1}{1.08^8} + \frac{0.6}{1.08^{6.5}} + \frac{0.4}{1.08^5} \right\} = 1.4\% 
\]

total 5.5% of pavement costs

Structure V \((\mu = 28 \text{ cm}, \sigma = 4 \text{ cm})\)

improvement: as structure II, 9.1%
Finally a comparison can be drawn between the costs calculated in this way for compensation of deviant layer thicknesses (in terms of \( \mu \) and \( \sigma \)) and the penalties determined on the basis of the penalty provisions (see table 8).

The penalty is then calculated on the basis of the mean layer thickness \( \bar{x} \) and the standard deviation \( s \) of the random sample. It should be noted that in this way the excess percentage of the penalty limits is calculated using estimates, calculated on the basis of a random sample, of the mean standard deviation \( (\bar{x} \text{ and } s) \).

Although these parameters must approximate as closely as possible to the true mean \( \mu \) and true standard deviation \( \sigma \), there may in fact always be a disparity between them.

The calculation is effected on the basis of the relationship:

Penalty (calculated as percentage of total pavement costs) = \( 0.2 \times \text{excess percentage in relation to penalty limit } B \), calculated from \( \frac{\bar{x} - R}{s} \)

the limit value \( R \) being fixed at 24.0 cm.

<table>
<thead>
<tr>
<th>structure</th>
<th>( \frac{\mu - R}{\sigma} )</th>
<th>%defectives = ( B )</th>
<th>penalty ( 0.2B ), % cost of pavement</th>
<th>calculated necessary compensation costs</th>
<th>cost/benefit ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal ( \frac{28 - 24}{2} = 2.0 )</td>
<td>2.3</td>
<td>zero*</td>
<td>zero</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>I ( \frac{27 - 24}{2} = 1.5 )</td>
<td>6.7</td>
<td>1.34</td>
<td>4.9</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>II ( \frac{26 - 24}{2} = 1.0 )</td>
<td>15.9</td>
<td>3.18</td>
<td>10.3</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>III ( \frac{24 - 24}{2} = 0 )</td>
<td>50</td>
<td>10</td>
<td>19.9</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>IV ( \frac{28 - 24}{3} = 1.33 )</td>
<td>9.2</td>
<td>1.84</td>
<td>5.5</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>V ( \frac{28 - 24}{4} = 1.0 )</td>
<td>15.9</td>
<td>3.18</td>
<td>11.4</td>
<td>3.2</td>
<td></td>
</tr>
</tbody>
</table>

* zero = no penalty because excess percentage is greater than a set limit value [8]
For details of this factor 0.2 and the penalty limit referred to here for the total layer thickness, reference should be made to the relevant comments on page 32 of this publication and to publication [8].

With the exception of the extreme case of structure III, a fairly constant relationship is thus found between the costs and 'estimates', i.e. on average 1 : 3. This means that the formula based on the deviation percentage for the total layer thickness is a good indication of the actual consequences of improvement of the pavement life.

In the relationship: penalty = factor \times \text{percent defectives}, the factor must, however, be at least 0.6 to obtain adequate compensation.

The dimension formula used above is a special instance of a general dimension formula derived as follows from the AASHO results:

\[
D = \Sigma \frac{c_i h_i}{0.004} = 0.004 \Sigma E_i^{\frac{3}{4}} h_i = \left( \frac{250}{E_s} \right)^{0.4} \left[ 1.158 \left\{ \frac{\Sigma n_i \left( \frac{P_i}{5} \right)^{4.2}}{0.107} \right\} - 1 \right]
\]

\(D\) = necessary layer thickness expressed as thickness index \(D\), up to \(p = 1.5\)

\(E_s\) = modulus of elasticity of subgrade

\(E_i\) = modulus of elasticity of pavement courses

\(h_i\) = layer thickness of pavement courses in cm

\(n_i\) = number of load repetitions of wheel load \(P_i\) during the service life (up to \(p = 1.5\)) of the road.

All in all, a safety coefficient of 0.11 \((D + \delta)\) is introduced on the basis of the 90\% reliability limits of the AASHO relationship and a combination with the (dynamic) modulus of elasticity \((= 100 \times \text{CBR})\) of the subgrade in the AASHO test \((\text{CBR} = 2.5)\).

In the first approximation, we have taken a proportionality of the value coefficient of a pavement material (of thickness \(h\) cm) with the modulus of elasticity to the power of 1/3.

For the purpose of comparative calculations, this formula is further simplified [11] to

\[D = K(CBR)^{-0.4} \left( \Sigma n_i P_i^{4} \right)^{\delta}\]

Although an infinite number of calculation examples can be worked out on the basis of this formula and on the basis of other design methods, a service life equation will always give broadly similar results, since these are based primarily on the low exponent of \(n\) while the subgrade characteristic which is introduced has a considerable influence on the absolute value of \(D\) but no influence at all on the relationship between the thickness indices for the same subgrade.
The parameter values chosen as an example are adapted to practical conditions and the use of a low determining CBR value (AASHO-CBR) in combination with reference to the topmost, well compacted layer of the subgrade, generally provides a sufficiently high safety coefficient to justify use of the original AASHO equation.

Finally it should be noted that the layer thickness variation of the total pavement has been treated in the foregoing as the determining factor for service life. Although variation both of the mean layer thickness and of the standard deviation has been taken into account, it seems reasonable in practice to take steps first of all to prevent excessive variation of the standard deviation. The provision that only the actually processed quantity can be invoiced – up to a maximum of 25 kg/cm per m$^2$ – is in itself a deterrent to excessively low average layer thicknesses.

Under normal circumstances processing of 'too little' asphalt will cut the contractor's profit.

It will also be expedient to lay down requirements for the thickness of the individual pavement layers, with reference to both the mean and standard deviation.

The greatest need for this applies to the topmost pavement layer and to the subbase, especially where this consists of a different material than the roadbase, e.g. sand cement.

For the top pavement layer a mean thickness of 4.3 cm has been calculated and a reasonable standard deviation is 0.5 cm.

With a mean layer thickness of 3.8 cm the total pavement thickness will be 0.5 cm lower, thus reducing the thickness index by 0.087 and leading to a reduction in service life of about 1.5 years in comparison with a normal pavement.

Expressed as a percentage of the cost of the surface asphalt layer, compensation for premature improvement will therefore be:

$$\frac{(1.08^{1.5} - 1)8}{1.08^{16}} + \frac{1}{1.08^{14.5}} \cdot \frac{100}{4} = 15\%$$

or 3%, expressed as a percentage of the total pavement costs.

On the basis of the choice of $R = 3.3$ cm, $\frac{\mu - R}{\sigma} = 1.0$ so that the excess percentage in this case is 15.9% and the penalty with a factor of 0.2 is equal to only 3.2% of the value of the asphalt layer or 0.6% of the cost of the asphalt pavement.

However, the performance of these calculations is far less logical in this case since a shortfall on the thickness of an underlying layer can be compensated by additional thickness of another layer, while a penalty is already imposed for shortfalls (although much greater in extent) on the (total) layer thickness. The above considerations show that there can be no question of an accumulated penalty exceeding the damage actually caused.
The main purpose of penalties in respect of the layer thickness of the surface layer is in any case to avoid excessive standard deviations so that the design calculations lend themselves far less to use as a basis for comparison. The harmful consequences of local shortfalls on thickness of the surface layers are sufficiently well known from practical experience.

For comparison, it should be noted that the deviation percentage of 15.9% calculated in this case will also be reached with a standard deviation of 1.0 cm if the layer thickness (as is usual in practice) coincides with the mean thickness. This means that the surface layer with a mean thickness of 4.3 cm (nominal thickness 4 cm) has a thickness of less than 3.0 cm over 10% of the road surface and of less than 2.5 cm over 3.6% of the surface.

**Voids**

A reduction in service life is the quantifiable expression of damage caused by excessive voids content. The calculation can be based on fatigue tests which have been carried out on asphalt bars by a number of research workers. In 1960, Saal and Pell [12] already described fatigue tests (constant $\sigma = 3\text{MN/m}^2$, 0°C, 50Hz) on sand asphalt bars (85% sand, 15% limestone filler, 91% (m/m) bitumen 40/50 M.E.). The voids content varied between 0 and 10% (V/V). They reported that the service life with 10% voids is only 1/16 of the life with 0% voids. The reduction is greater than would be expected from the lower quantity of asphalt in the cross-section so that the cause must lie in stress concentration.

Bazin and Saunier [13] also performed fatigue tests on a sand asphalt of the same composition (constant $\sigma$, 10°C, 50 Hz) with mean voids of 4.5, 7.0 and 9.0%. Their test results ($\varepsilon - n$ at failure graph) show for example that with 4.5% voids the number of load repetitions ($n$) to failure is $10^6$ with $\varepsilon = 200 \times 10^{-6}$; this number is reduced to $3 \times 10^5$ with 9% voids. A 4.5% increase in voids thus reduces the service life on average to one third of its initial value. There is also a reduction in rigidity (approx. 3% - relative - per percent voids) so that the loss of service life for an identical stress level ($\sigma - n$ to failure graph) will be even greater.

Monismith and co-workers [14] carried out large scale fatigue tests in which, as in the research referred to above, the voids varied on the basis of compaction differences, in other words the degree of compaction of identical asphalt varied.

This research was conducted with two mixes of different composition: a BS 594 hot rolled asphalt with 30% chippings and 7.9% bitumen 40/50, and two Californian asphaltic concrete mixes with on average 55% chippings and approximately 6% bitumen (85/100). The Wöhler curves (relationship between strain at failure and number of load repetitions at which failure occurs, in the general form $\varepsilon = A \cdot n^{-b}$) for the two mixes show considerable differences, in particular a substantially higher
strain at failure and a flatter curve (higher exponent $b$) for BS 594 asphalt. This better fatigue characteristic was also reflected in behaviour with increasing voids: a 10% increase in voids for BS 594 asphalt and 4-6% increase for Californian asphalitic concrete caused a reduction in service life with a factor of 10. Since the voids content of the latter mix was distributed over larger voids, it was concluded that this voids distribution is the cause of the difference.

Rigidity is reduced as follows by the increase in voids: in the case of BS 594 asphalt with 5% voids by a factor of $2/3$ and 11% voids by a factor of $1/3$; in the case of Californian asphalt for a 5% increase in voids, on average by a factor of $1/2$.

The reduction in service life cannot, however, be fully explained by this loss of rigidity.

In recent years, Pell and Taylor [15] have conducted extensive series of fatigue tests and established a complete $\varepsilon - n$ graph for voids varying in 1% stages between 0 and 10%. The results relate to BS 594 roadbase asphalt, i.e. discontinuously graded asphalitic concrete with 60% porphyry chippings, round sand or crushed porphyry sand, limestone or sand filler and 6% bitumen 40/50, investigated with constant $\sigma(1.2$ MN/m$^2$), at 20°C and 20Hz.

The description of the research shows, however, that these results were not obtained by intentional variation of the degree of compaction but that they are primarily due to differences in mix composition as the origin of the varying voids (i.e. variations in bitumen and filler contents since other variations were not significant). The results show that an increase in voids from 0 to 10% gives a reduction in service life by a factor of 1/20 and that a 1% voids increase signifies a multiplication factor of slightly less than 3/4 for the service life (reduction of about 25% in service life).

J. M. Kirk [16] also performed a number of fatigue test (constant $\varepsilon$, 50 Hz) mainly with sand asphalt (e.g. based on crushed granite sand, limestone filler and 80/100 bitumen) but also with a number of very fine asphalitic concrete mixes with variable voids (1-18%). For these mixes, he reports the strain at failure at $10^6$ load repetitions: for other mixes with high voids the strain at failure is relatively low, but these mixes also have a lower bitumen content to which the difference is attributed. Three mixes with the same bitumen content but differing by 6% in voids, show practically the same value for $\varepsilon$ at $n = 10^6$: Kirk accordingly concludes that voids do not influence the fatigue strain.

The conclusion reached by Kirk differs completely from the results of the other research workers but is reported here for completeness. General considerations of material engineering also tend to throw doubt on Kirk's result.

The other research workers referred to above accordingly show a varying influence of the increase in voids on fatigue life. To sum up, the results in respect of a 10%
increase in voids due entirely to variations in the degree of compaction, show a reduction factor for service life of
\[
\frac{1}{16} \quad \text{(Saal and Pell, sand asphalt)}
\]
\[
\frac{1}{10} \quad \text{(Bazin and Saunier, sand asphalt)}
\]
\[
\frac{1}{100} \quad \text{(Epps and Monismith, Californian asphaltic concrete), and}
\]
\[
\frac{1}{10} \quad \text{(Epps and Monismith, BS 594 hot rolled asphalt)}
\]
and primarily due to mix variations
\[
\frac{1}{20} \quad \text{(Pell and Taylor, asphaltic concrete)}.
\]
In practice, however, the variation in voids is a combination of variation in the degree of compaction and mix; in some cases one of these two factors predominates.

Having regard to the fact that the asphaltic concrete used in the Netherlands is more closely related to American ‘Marshall asphalt’ than to English ‘hot rolled asphalt’ and that its composition, under the influence of various other factors, is (unfortunately) increasingly tending to resemble that of Marshall asphalt, the assumption of a reduction in service life of 25% (relative) per volume per cent increase in voids would seem to be rather on the low side (factor 1/18 for 10% voids).

It is self-evident that the differences between the results obtained may be attributable to the widely varying test methods used by the different authors, but this aspect has not been studied in more detail.

A number of further observations are called for on the use made below of the fatigue life determined in laboratory tests of asphalt. It must not be assumed that the laboratory fatigue strength or strain is the decisive value in practice but only that the relative reduction determined in the laboratory will also be reflected in practice; furthermore the road design should be considered to be based on a fatigue strength which will be just adequate at the end of the service life of the road. It is accordingly assumed that the strain of the asphalt is in practice the decisive criterion for service life and that reduction in service life under practical conditions will be proportional to the laboratory situation.

This assumption will not always hold good, e.g. because during construction in accordance with the design data another criterion, soil pressure which will be dealt with later, may be the principal determinant, or because the location of the asphalt layer is such that it is subject to less strain. It must, however, be noted that, partly because of the reduced strain at failure of the asphalt, this characteristic which is contrary to the design data may become the decisive factor (and indeed often does as is shown by Brouwers [11]), while in a ‘protected’ asphalt layer the strain which occurs may increase with the passage of time so that this weakness criterion again applies. In the event of failure of the underlying layer as a result of an effective reduction in layer thickness even higher requirements may be placed on this material.

The second assumption can be allowed for in practice with rational design methods, in which the strain at failure calculated through laboratory fatigue tests is taken as the failure criterion.
So as not to make these calculations excessively complicated, it is assumed that the rigidity relationship of the total asphalt pavement (constructed out of several different layers) to the roadbase or subgrade is not changed; allowance is thus only made for changes in the failure criterion (determinant strain). This of course means that a further consequence, a change in the strain of the pavement occurring under load, is disregarded.

If this factor is to be taken into account, the structure will have to be the subject of further calculation, e.g. with the aid of the BISTRO programme. A simple calculation is given here on the basis of which it can be seen that the strain which occurs with increased voids falls further: working from the calculation method described by Burmister (Fox, Acum and Fox), and Jeuffroy, Brouwers [17] concluded that the following expression is reasonably accurate:

\[ \log\frac{\sigma_r}{\sigma_o} = A \log \frac{E_2}{E_s} + B, \]

with which parallel lines are found for different values of \( \frac{E_1}{E_2} \).

Here \( \sigma_r \) is the radial strain at the bottom of the asphalt layer
\( \sigma_o \) is the contact pressure
\( E_1, E_2 \) and \( E_s \) are the respective moduli of elasticity of asphalt, roadbase and subgrade (applicable to 2 and 3 layer systems with \( \gamma = 0.5 \)).

The following expression [18] is a good approximation for the 2 layer system:

\[ \frac{\sigma_r}{\sigma_o} = 1.8 \left( \frac{a}{h} \right)^{1.56} \left( \frac{E}{E_s} \right)^{0.15} \]

applicable for Jeuffroy and Burmister if \( E/E_s > 20 \),
where \( h \) is the layer thickness
\( a \) is the radius of the contact surface
\( E \) is the modulus of elasticity of the pavement (rigidity)
\( E_s \) is the modulus of elasticity of the subgrade.

If the modulus of elasticity of the asphalt pavement is halved (according to Monismith this is the consequence of 5% higher voids in Californian asphaltic concrete), the tensile stress encountered at the bottom of the asphalt layer is only reduced by 10%. As a consequence of the halving of the modulus of elasticity, the strain which occurs will clearly be substantially greater. In disregarding this factor one is therefore postulating a far more favourable situation than exists in reality.

On the basis of the assumed 25% (relative) reduction in service life for each 1 per cent higher voids content, the financial consequences for the road authority and contractor must now be calculated.
The method of calculation is similar to that used for the layer thickness. This is done on the basis of the average situation for bitumen-bound gravel obtaining in the past, i.e. the ‘average’ pavement with mean voids of 5.9% and a standard deviation ($s$) of 1.8% is taken as standard and assumed to have a service life of 20 years. A ‘cost-benefit’ calculation is now effected for four different situations with higher average voids; on the one hand the financial compensation for the resulting reduction in service life is calculated, while on the other the penalty is determined with reference to the ‘equivalent’ statistical penalty provisions (see Table 9). The compensation is calculated in the same way as for layer thickness (with the costs of the asphalt layer set at $14$ per m$^2$). It is assumed that the true standard deviation ($\sigma$) is and remains identical to the ‘historical’ mean value of 1.8%. Similar calculations based on different standard deviations are of course also possible but have not been carried out here. Consequently the calculation is not as complete as for the layer thickness, where allowance was also made for the consequences of increased maintenance due to dispersion in the layer thickness.

It is thus apparent that the proceeds of the penalty represent on average only 15% of the calculated costs to the road authority.

<table>
<thead>
<tr>
<th>mean voids (vol. %)</th>
<th>mean life (years)</th>
<th>compensatory asphalt layer at end of service life, % value</th>
<th>proceeds of penalty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>asphalt layer</td>
<td>($R = 9.5%$) % defectives penalty</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>($s = 1.8%$) $B$ as % value</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\frac{R - \mu}{\sigma}$ of asphalt layer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(K = 0.38)$</td>
</tr>
<tr>
<td>5.9</td>
<td>20</td>
<td>zero</td>
<td>9.5 - 5.9 = 2.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\frac{(1.08^3 - 1)8}{1.08^{20} + \frac{1}{1.07^{17}}} \frac{1}{4} = 18$</td>
<td>2.3 zero</td>
</tr>
<tr>
<td>6.9</td>
<td>17</td>
<td>$\frac{(1.08^6 - 1)8}{1.08^{20} + \frac{2}{1.08^{14}}} \frac{1}{4} = 42$</td>
<td>1.44 7.5 2.3</td>
</tr>
<tr>
<td>7.9</td>
<td>14.4</td>
<td>$\frac{(1.08^8 - 1)8}{1.08^{20} + \frac{3}{1.08^{12}}} \frac{1}{4} = 66$</td>
<td>0.89 18.7 5.6</td>
</tr>
<tr>
<td>8.9</td>
<td>11.2</td>
<td>$\frac{(1.08^{10} - 1)8}{1.08^{20} + \frac{4}{1.08^{10}}} \frac{1}{4} = 96$</td>
<td>0.33 37.0 11.1</td>
</tr>
<tr>
<td>9.5</td>
<td>10</td>
<td>$\frac{1}{1.08^{20} + \frac{4}{1.08^{10}}} \frac{1}{4} = 93$ based on add. treatment excl. compensation for loss of use</td>
<td></td>
</tr>
</tbody>
</table>

47
These calculations show the same overall pattern for open-textured and dense asphaltic concrete (see Table 10).

It should be noted that the values for $\mu$ and $\sigma$ chosen for open-textured and dense asphaltic concrete are the average values for mean voids and mean standard deviation found by analysis of all projects on which penalties have been imposed hitherto [8].

The value chosen for the limit value $R$ is also based on the proposed new statistical procedure for the determination of penalties [8].

The limit value used at present for penalty calculations is $R = 9.5\%$ for open-textured asphaltic concrete and $R = 6.5\%$ for dense asphaltic concrete.

Where these limit values are used, the results shown in brackets should be obtained. In that case there is a much greater discrepancy between the calculated penalty and the costs calculated on the basis of the reduction in service life for open-textured asphaltic concrete. In the case of dense asphaltic concrete the difference is less marked, corresponding to a factor of only $4\frac{1}{2}$ instead of 6.

When open-textured asphaltic concrete is applied as basecourse on a sufficiently rigid road base, a reduction in the permissible strain may be less applicable: however the consequences of the reduction in rigidity will still apply.

In the foregoing, only the consequences of a reduction in the strain at failure of asphalt as a result of an increase in voids have been considered and the increase in the strain attributable to a reduction of rigidity has been disregarded.

Table 10  Penalties for open-textured and dense asphaltic concrete due to higher voids content

<table>
<thead>
<tr>
<th>mean voids (vol. %)</th>
<th>$\frac{R - \mu}{\sigma}$</th>
<th>percentage defectsives $B$</th>
<th>penalty as % value of asphalt layer $K = 0.3B$</th>
<th>mean voids (vol. %)</th>
<th>$\frac{R - \mu}{\sigma}$</th>
<th>excess percentage defectsives $B$</th>
<th>penalty as % value of asphalt layer $K = 0.3B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7</td>
<td>2.00</td>
<td>2.3</td>
<td>zero</td>
<td>3.7</td>
<td>2.00</td>
<td>2.3</td>
<td>zero</td>
</tr>
<tr>
<td>mean standard</td>
<td>2.00</td>
<td>2.3</td>
<td>zero</td>
<td>3.7</td>
<td>2.00</td>
<td>2.3</td>
<td>zero</td>
</tr>
<tr>
<td>5.7</td>
<td>1.47</td>
<td>7.1</td>
<td>2.1</td>
<td>4.7</td>
<td>1.39</td>
<td>8.2</td>
<td>2.5</td>
</tr>
<tr>
<td>(2.00)</td>
<td>(2.3)</td>
<td>(zero)</td>
<td>(2.1)</td>
<td>(2.09)</td>
<td>(13.8)</td>
<td>(4.1)</td>
<td></td>
</tr>
<tr>
<td>6.7</td>
<td>0.95</td>
<td>17.1</td>
<td>5.1</td>
<td>5.7</td>
<td>0.79</td>
<td>21.5</td>
<td>6.5</td>
</tr>
<tr>
<td>(1.47)</td>
<td>(7.1)</td>
<td>(2.1)</td>
<td>(2.09)</td>
<td>(31.2)</td>
<td>(9.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.7</td>
<td>0.42</td>
<td>33.7</td>
<td>10.1</td>
<td>6.7</td>
<td>0.18</td>
<td>42.9</td>
<td>12.9</td>
</tr>
<tr>
<td>(0.95)</td>
<td>(17.1)</td>
<td>(5.1)</td>
<td>(5.12)</td>
<td>(54.8)</td>
<td>(16.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.5</td>
<td>0</td>
<td>50</td>
<td>15.0</td>
<td>7.0</td>
<td>0</td>
<td>50</td>
<td>15.0</td>
</tr>
<tr>
<td>(0.53)</td>
<td>(29.8)</td>
<td>(8.9)</td>
<td>(0.30)</td>
<td>(61.8)</td>
<td>(18.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
However, a further consequence of this loss of rigidity which can easily be determined in numerical terms must now be considered, i.e. the increase in the occurring soil pressure.

As we have already seen, according to Epps and Monismith the loss of rigidity for two widely different asphalt mixes is represented by a factor of 1/2 and 2/3 respectively for 5% voids, corresponding to a reduction of 8 and 13% respectively per percent voids.

11% is taken as the mean value here, i.e. a reduction factor of 0.89.

The relationship between the soil pressure $\sigma_v$ (directly below the pavement in the centre of the load) can generally be indicated as follows according to Brouwers [17] for the 2 and 3 layer system used by Burmister and Jeuffroy:

$$\log \frac{\sigma_v^2}{\sigma_o} = -A \log \frac{E_2}{E_3} - B$$

in which $A$ is practically always equal to 2/3, and

$$B = f \left( \frac{a}{h_1}, \frac{a}{h_2}, \frac{E_1}{E_2} \right)$$

The following expression is arrived at for the 2 layer system [18]:

$$\sigma_v = 2\sigma_o \left( \frac{a}{h} \right)^2 \left( \frac{E_s}{E} \right)^{2/3}$$

applicable to Burmister and Jeuffroy's system where $\frac{E}{E_s} \geq 40$.

In the complete asphalt structure, considered for greater simplicity as a 2 layer system, $\sigma_v$ is thus inversely proportional to the rigidity of the asphalt to a power of 2/3. This means that a reduction in $E$ by a factor of 0.89 results in a reduction of $\sigma_v$ by a factor of 1.08.

On the basis of the AASHO test results a relationship of the type $\sigma_v = A n^{-5}$ has been established between the service life expressed in axle passages of an equivalent axle ($n$) and the soil pressure ($\sigma_v$); this expression has been used as the basis for the Shell design method [19].

This means that an increase of $\sigma_v$ by a factor of 1.08 leads to a reduction of $n$ by a factor of 0.68, i.e. approx. 2/3.

This result is even more unfavourable than the factor of 3/4 calculated for the reduction in strain at failure.

These factors obviously have consequences if the soil pressure is or becomes decisive for the pavement as is assumed in the Shell design method, e.g. for thick asphalt structures (of good quality in terms of strain at failure).

Comparative calculations ('cost-benefit analyses') have been made on the basis of this $2/3$ factor which reduces the service life (Table 11).

In drawing the comparison between the costs of compensating asphalt and the yield of the penalty, it has been assumed here (for greater simplicity) that the entire struc-
Table 11 Consequences of reduced rigidity of the entire structure due to higher voids content

<table>
<thead>
<tr>
<th>increase in voids above standard, all layers, %</th>
<th>mean service life in years</th>
<th>cost of compensatory asphalt layer at end of service life, % value of total pavement</th>
<th>proceeds of penalty as % value of total pavement, mean of 5 layers (see Tables 6 and 7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20</td>
<td>zero</td>
<td>zero</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>[ \frac{(1.08^4 - 1)8}{1.08^{20}} + \frac{2}{1.08^{16}} ] 100 \times \frac{20}{20} = 6 ]</td>
<td>[ 3 \times 2.3 + 2.1 + 2.5 ] \times \frac{5}{5} = 2.3</td>
</tr>
<tr>
<td>2</td>
<td>12\frac{1}{2}</td>
<td>[ \frac{(1.08^8 - 1)8}{1.08^{20}} + \frac{3}{1.08^{12}} ] 100 \times \frac{20}{20} = 13 ]</td>
<td>[ 3 \times 5.6 + 5.1 + 6.5 ] \times \frac{5}{5} = 5.7</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>[ \frac{(1.08^{11} - 1)8}{1.08^{20}} + \frac{4}{1.08^{9}} ] 100 \times \frac{20}{20} = 21 ]</td>
<td>[ 3 \times 11.1 + 10.1 + 13.3 ] \times \frac{5}{5} = 11.3</td>
</tr>
<tr>
<td>or [ \frac{8}{1.08^9} \times \frac{100}{4} = 20 ]</td>
<td>based on additional treatment, excluding compensation for loss of use</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ture (i.e. all the layers) will have a given higher voids content (as compared with the mean standard voids) so that the ‘yield’ and the costs are expressed as a percentage of the cost of the entire pavement.

In these calculations too there is a discrepancy between the compensation costs to be fixed on the basis of the deterioration of material characteristics and the yield of the penalty imposed under the relevant provisions. Finally it should be noted that in practice it will generally not be possible to determine as such the possible loss of quality or damage resulting from the material-technical phenomena described here. It is also not really known to what extent a safety margin is included under practical conditions, as a result of which deviations from the mean (which is treated under the penalty provisions as a feasible aim) during the normal service life (which may in fact also be limited by other factors) will not lead to manifest increases in voids. Should this be so, it would still not disprove the value of the cost-compensation method used. The relationship between the voids and all the ‘strength characteristics’ of the material, and the interrelationship between the latter and the design thickness, make it reasonable to suppose that this particular property is of considerable importance to the behaviour pattern, especially and above all as regards the durability (service life) of the structure.

It will perhaps not be superfluous to point out that in this study various other consequences of higher voids which also influence durability (mechanical deterioration and stripping of the binder under the influence of climate and traffic loads, especially in the case of the surface layers) have been entirely disregarded.

50
Bitumen content

The consequences of deviations in the asphaltic bitumen contents can be calculated in the same way as has been described in detail for the voids. Saal and Pell [12] and Jimenez and Gallaway [20] have already indicated the great influence of the asphaltic bitumen content on the service life in fatigue tests of asphalt. In their first fatigue tests, Saal and Pell already gained the impression that the permissible strain in the binder \( \varepsilon_B \) is the determining parameter. The strain in the mix is then \( \varepsilon_m = B_v \cdot \varepsilon_B \) where \( B_v \) is the volumetric concentration of the bitumen in the mix. Jimenez and Gallaway found an optimum binder content at which the service life is greatest for a given asphalt mix.

The first finding has been confirmed by other research workers (naturally below the optimal binder content), e.g. by Kirk [16] \((8 \cdot 10^{-6} \text{ per } \% (V/V) \text{ bitumen at } n = 10^6)\) and Epps and Monismith [14], while an optimum for a curve giving the service life as a function of the binder content was later also found by Pell [21], Pell and Taylor [15] and Epps and Monismith [14].

Pell's assumption that the relationship between \( n \) and \( \varepsilon_B \) is decisive, with \( n = A \cdot \varepsilon_B^{-5} \), gives, in conjunction with \( \varepsilon_M = B_v \cdot \varepsilon_B : n = A \cdot B_v^{-5} \cdot \varepsilon_M^{-5} \text{ or } \varepsilon_M = A' \cdot B_v^{5} \cdot n^{-\frac{1}{5}} \). This means that for \( n = \text{constant} \) the permissible strain in the mix is proportional to the volumetric concentration of the bitumen, while for an identical strain level the permissible number of load repetitions (the service life) is inversely proportional to \( B_v^{-5} \), i.e.

\[
{n = n_0 \left( \frac{B_v}{B_{v_0}} \right)^5}
\]

This means for example that in a comparison between an asphalt with 7.0\% (m/m) and an asphalt with 6.3\% bitumen (10\% less) – with (given 0\% voids and a mineral density of 2.65) \( B_v = 0.15 \) and 0.14 respectively (ratio of 0.93) – the service life in the case of 6.3\% binder will be 0.7 times the life with 7.0\% bitumen.

This overall approach to the problem has in fact been superseded by the practical determination of the service life of a number of asphalt mixes as a function of the binder content. Pell's study [21] related in the first instance to a BS 594 roadbase asphalt (discontinuously graded) with 60\% porphyry chippings and sand to which, in accordance with the relevant specifications, no filler is added, and 4-12.5\% bitumen 40/50. Under Pell's test conditions \((\sigma_{\text{max}} = 1/2 \text{ N/mm}^2, 10^\circ \text{C, 20 Hz})\) a definite maximum service life of \( n = 1.5 \cdot 10^5 \) was found for 7.7\% bitumen. His graph for \( n = f \) (bitumen \% ) shows that \( n \geq 10^5 \) between 6.5 and 10\% bitumen, while \( n \) falls much more rapidly with lower than with higher binder contents (\( n = 3.10^3 \) for 4\% bitumen).
The voids varied between 0 and 9.8% and corresponded to 0.9% in the case of the optimum mix. This study was followed by a further study in which for mixes with the desired binder content of 6%, according to BS 594, the filler content (limestone) was varied between 0 and 17%. Here again a definite optimum service life was found for 9.7% filler; the original value \( n = 8.10^4 \) (without filler) was increased in this case to \( 2.5 \cdot 10^6 \). The voids ranged from 0 to 4.5% (0 in the optimum case). Finally Pell and Taylor [8] kept a constant filler content (10.3%) but varied the binder content between 3.5 and 10.5%; it was then found that 6.4% bitumen 40/50 gave an optimal service life \( (2.5 \cdot 10^6) \) but this fell off quickly, especially with lower binder contents. Even at the limits of the BS 594 specification, i.e. 5.9% min. and 7.1% max., a considerable reduction is found, i.e. to \( 4 \cdot 10^4 \) and \( 1.05 \cdot 10^5 \). The voids varied between 0 and 5.4%.

In Pell’s first study, the optimum mix with 8.3% bitumen also showed a maximum rigidity of \( 7.5 \cdot 10^3 \) N/mm² (probably because with a lower bitumen content the increase is counteracted by the higher voids content; in Pell and Taylor’s series a mix with a lower binder content had a higher rigidity than the optimum mix \((11 \cdot 10^3 \text{ MN/m}^2)\). The decisive strain level in these tests was 100 to 150 \( \cdot 10^{-6} \). A further striking feature is that the optimum binder content is 6.3% with 10.3% filler, but substantially greater, i.e. 8.3%, without filler.

The fact that specific mix characteristics may influence the optimum bitumen content and the corresponding service life was already observed by Jimenez and Gallaway, who found in particular that a rougher, more absorbent aggregate had a higher optimum bitumen content and a longer service life. Pell and Taylor also compared a sand filler with limestone which gave lower results.

Epps and Monismith also studied the influence of the bitumen content for a number of mixes. This was done first with two asphaltic concrete mixes on the basis of two types of basalt aggregate (with approx. 50% chippings, max. 12 mm, 6% filler and 5.3-8.7% bitumen 60/70). Under his test conditions \((\sigma_{\text{max}} = 1.1 \text{ MN/m}^2, 20^\circ \text{C}, 0.1 \text{ s and a strain level of approx. 500 } \mu \text{m})\), Monismith found a maximum \( n = 1.9 \cdot 10^4 \) at 6.7% bitumen. As in Pell’s study, \( n \) drops very rapidly with lower asphaltic bitumen content and less rapidly with a higher content (still \( 10^4 \) at 8.5%). A maximum rigidity is again found \((2.7 \cdot 10^3 \text{ MN/m}^2)\) at the optimum binder content. The voids in this case were 5.2% and varied between 1.6 and 8.8% in other instances. The optimum binder content for maximum service life at 6.7% was 0.8% higher than on the basis of the Californian (stability) tests (cohesiometer) while the service life is then roughly three times greater.

Analysis of the results obtained by Pell and Taylor for the mix with 10% limestone filler (far more comparable with our own asphalt mixes than the mix without filler) shows that the life of \( 2.47 \cdot 10^6 \) at 6.4% bitumen content is reduced to \( 4.43 \cdot 10^5 \) with 4.5% bitumen, and to \( 1.05 \cdot 10^4 \) with 3.5% bitumen. However, with these mixes
the voids rise from 0.4 to 2.8 or 5.4% so that (on the basis of the data indicated in the chapter on voids) a halving of the service life can be shown for each successive bitumen content. The reduction factor resulting from the respective lowering of the bitumen content by 1.9 and 1.0% is then 0.36 and 0.05 respectively which corresponds to a reduction factor of 0.90 and 0.55 respectively per 0.2% bitumen. This reduction therefore increases considerably with a very low bitumen content (departing from the general formulation based on $B_e$).

Epps and Monismith describe a reduction from $n = 19.1 \cdot 10^3$ for 6.7% bitumen and 5.2% voids to $n = 6.6 \cdot 10^3$ at 6.2% bitumen and 6.2% voids. After correction for the considerable drop in $n$ found by them at higher voids content (factor 0.63 per % voids) we obtain a reduction factor of 0.55 for 0.5% bitumen or 0.80 for 0.2% bitumen. For an even lower binder content, Epps and Monismith describe a less far-reaching reduction in $n$. Further analysis of their results shows, however, that this reduction can also be attributed to the further increase in voids content. It is therefore considered reasonable in our further discussion to take a mean reduction factor for service life of 0.85 per 0.2% bitumen applicable to binder contents which are not reduced too far.

For raised bitumen contents the reduction factor is lower, i.e. a factor of approximately 0.6 per % bitumen.

The results of calculations to which the same assumptions apply as for voids are shown in Table 12.

The calculation is effected for coarse, dense asphaltic concrete and, in addition to the assumed mean standard deviation of 0.29% (identical to the historical average for all penalty contracts), a calculation has also been made of the penalty for an unnecessarily high standard deviation of 0.35%.

As in the case of voids we did not investigate the effect of higher standard deviations on the increase in maintenance costs.

Similar calculations can be made for other asphalt layers. The results obtained here show that the yield of the penalty is always substantially lower than the damage actually incurred.

For higher bitumen contents the results will be less unfavourable, but it must be remembered that with high binder contents other types of damage which are difficult to quantify are likely to occur.

If the mean bitumen content is reduced a quantity of binder the cost of which can be calculated will not be supplied. In the examples quoted with 6.3, 6.1 and 5.85% bitumen respectively, the value of this non-delivery is not less than 0.6%, 1.2% and 1.9% (overall) respectively of the value of the asphalt (on the basis of prices obtaining in mid-1972). This percentage must be deducted from the proceeds of the penalty if the comparison is to be effected in the same manner as in respect of the other characteristics.
Table 12  Coarse dense asphaltic concrete: consequences of reduced bitumen content

<table>
<thead>
<tr>
<th>mean bitumen content %</th>
<th>mean service life (years)</th>
<th>compensatory asphalt layer at end of service life as % value of asphalt layer</th>
<th>proceeds of penalty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$R = 5.85%$</td>
<td>$s = 0.28%$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$R - \mu$</td>
<td>$s - \sigma$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma$</td>
<td>penalty as %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>per cent defectsives $B$</td>
<td>value asphalt layer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$K = 0.3B$</td>
<td></td>
</tr>
</tbody>
</table>

| 6.5 | 20 | zero | $6.5 - 5.85$ | $0.29$ | $2.24$ | 1.25 | none |
| 6.3 | 18½ (18) | $\left\{ \frac{(1.08^3 - 1)8}{1.08^{20}} + \frac{1}{1.08^{18}} \right\} \frac{100}{4} = 13$ | $6.3 - 5.85$ | $0.29$ | $1.86$ | 3.1 | none |
| 6.1 | 16½ (17) | $\left\{ \frac{(1.08^3 - 1)8}{1.08^{20}} + \frac{1}{1.08^{17}} \right\} \frac{100}{4} = 18$ | $6.1 - 5.85$ | $0.29$ | $1.29$ | 9.9 | 3.0 |
| 5.85 | 15 | $\left\{ \frac{(1.08^3 - 1)8}{1.08^{20}} + \frac{2}{1.08^{15}} \right\} \frac{100}{4} = 36$ | $5.85 - 5.85$ | $0.35$ | $0.71$ | 23.9 | 7.2 |

Conclusion

The penalties calculated with the aid of the new statistical method of evaluation amount on average to 30, 15 and 30% respectively of the necessary compensation costs which can be calculated on the basis of the reduction in the service life of an asphalt pavement as a consequence of certain shortcomings in respect of layer thickness, voids and bitumen content.
A uniform, effective and equitable system of quality control for pavements has been applied to road-building projects in the Netherlands since 1968. The system has been developed in consultation between the public authorities and the contractors' organizations.

The system is characterized by a distinction between regular, daily production control by the contractor and limited retrospective acceptance control by the public authority based on random samples.

Under the existing non-statistical system, samples (one per 2000 m$^2$) are examined to determine the layer thicknesses, sand cement strength and density and bitumen content of the asphalt.

When the quality does not meet the specified values, financial penalties are applied the level of which depends on the extent of the deviation from the set standards. In the light of experience with over 300 major projects, the system has now been developed into an almost entirely statistical method of control which is ready for practical introduction.

Under this statistical system, samples are examined for road surface areas up to a maximum of 200,000 m$^2$ ($n = 20$ or $40$). Penalties are imposed if $\frac{|R - \bar{x}|}{s} < Q$, where $R$ is the quality limit, $\bar{x}$ the mean value, $s$ the standard deviation and $Q$ the quality index ($Q = 1.6$ or $1.4$).

Penalties are fixed by the formula $K = 0.3B - C$, where $K$ is the penalty as a percentage of the pavement costs, $B$ the percentage of unsatisfactory work and $C = 1.0$ or $2.0$.

Control of the surface characteristics relates to skid-resistance and evenness measured at random points on 30% of the pavement surface.

Penalties are imposed if the measurement results do not meet the criteria. If the results fall below certain safety limits the surfaces must be repaired at the contractor's cost.

A theoretical study of the shortening of the service life of a road pavement due to deviations in the thickness, density and bitumen content of asphalt layers, leads to the conclusion that the penalties imposed under the statistical quality control system are substantially lower (on average three times lower) than the calculated costs theoretically necessary as compensation for the reduction in value due to the shorter service life. Calculations of this kind are set out in the final chapter. The reduction in the service life of the pavement as a consequence of specific shortfalls on layer thickness, excessive voids and too low bitumen content is calculated. The consequences of changes in the service life pattern in relation to reconstruction and maintenance,
in addition to the costs of the resulting necessary additional work in comparison with the specified work structure, are also calculated.

Finally, the amount needed after laying the pavement to cover the costs calculated in this way is also determined to give the necessary compensation costs. These amounts are compared with the proceeds of the penalties imposed in such cases.
Literature


In the series of Rijkswaterstaat Communications the following numbers have been published before:

No. 1.* Tidal Computations in Shallow Water
   Dr. J. J. Dronkers† and prof. dr. ir. J. C. Schönfeld
   Report on Hydrostatic Levelling across the Westerschelde
   Ir. A. Waalewijn, 1959

No. 2.* Computation of the Decca Pattern for the Netherlands Delta Works
   Ir. H. Ph. van der Schaaff† and P. Vetterli, Ing. Dipl. E. T. H., 1960

No. 3.* The Aging of Asphaltic Bitumen

No. 4. Mud Distribution and Land Reclamation in the Eastern Wadden Shallows
   Dr. L. F. Kamps†, 1962

No. 5. Modern Construction of Wing-Gates
   Ir. J. C. le Nobel, 1964

No. 6. A Structure Plan for the Southern IJsselmeerpolders
   Board of the Zuyder Zee Works, 1964

No. 7. The Use of Explosives for Clearing Ice
   Ir. J. van der Kley, 1965

No. 8.* The Design and Construction of the Van Brienenoord Bridge across the River Nieuwe Maas
   Ir. W. J. van der Ebit, 1968

No. 9. Electronic Computation of Water Levels in Rivers during High Discharges
   Section River Studies. Directie Bovenrivieren of Rijkswaterstaat, 1969

No. 10.* The Canalization of the Lower Rhine
   Ir. A. C. de Gaay and ir. P. Blokland†, 1970

No. 11.* The Haringvliet Sluices
   Ir. H. A. Ferguson, ir. P. Blokland† and ir. drs. H. Kuiper, 1970

No. 12. The Application of Piecewise Polynomials to Problems of Curve and Surface Approximation
   Dr. Kurt Kubik, 1971

No. 13. Systems for Automatic Computation and Plotting of Position Fixing Patterns
   Ir. H. Ph. van der Schaff†, 1972

No. 14. The Realization and Function of the Nothen Basin of the Delta Project
   Deltadienst of Rijkswaterstaat, 1973

No. 15. Fysical-Engineering Model of Reinforced Concrete Frames in Compression
   Ir. J. Blaauwendraad, 1973

No. 16. Navigation Locks for Push Tows
   Ir. C. Kooman, 1973

No. 17.* Pneumatic Barriers to reduce Salt Intrusion through Locks
   Dr. ir. G. Abraham, ir. P. van der Burg and ir. P. de Vos, 1973

* out of print
In the series of Rijkswaterstaat Communications the following numbers have been published before (continuation):

No. 18. Experiences with Mathematical Models used for Water Quality and Water Problems
Ir. J. Voogt and dr. ir. C. B. Vreugdenhil, 1974

No. 19.* Sand Stabilization and Dune Building
Dr. M. J. Adriani and dr. J. H. J. Terwindt, 1974

No. 20. The Road-Picture as a Touchstone for the three dimensional Design of Roads
Ir. J. F. Springer and ir. K. E. Huizinga (also in German), 1975

No. 21. Push Tows in Canals
Ir. J. Koster, 1975

No. 22. Lock Capacity and Traffic Resistance of Locks
Ir. C. Kooman and P. A. de Bruijn, 1975

No. 23. Computer Calculations of a Complex Steel Bridge verified by Model Investigations
Ir. Th. H. Kayser and ir. J. Binkhorst, 1975

No. 24. The Kreekrak Locks on the Scheldt-Rhine Connection
Ir. P. A. Kolkman and ir. J. C. Slagter, 1976

No. 25. Motorway Tunnels built by the Immersed Tube Method
Ir. A. Glerum, ir. B. P. Rigter, ir. W. D. Eysink and W. F. Heins, 1976

No. 26. Salt Distribution in Estuaries
Rijkswaterstaat, Delft University of Technology, Delft Hydraulics Laboratory, 1976

No. 27. Asphalt Revetment of Dyke Slopes
Committee on the Compaction of Asphalt Revetments of Dyke Slopes, 1977

No. 28. Calculation Methods for Two-dimensional Groundwater Flow
Dr. ir. P. van der Veer, 1978

* out of print