A STUDY OF STANDARD SKIES CLASSIFICATION

TANG HO LUN

MASTER OF PHILOSOPHY

CITY UNIVERSITY OF HONG KONG

DECEMBER 2008

CITY UNIVERSITY OF HONG KONG
香港城市大學

A STUDY OF STANDARD SKIES CLASSIFICATION
標準天空分類之研究

Submitted to
Department of Building and Construction
建築學系
in Partial Fulfillment of the Requirements
for the Degree of Master of Philosophy
哲學碩士學位

by

Tang Ho Lun
鄧浩麟

December 2008
二零零八年十二月

ABSTRACT

In 2003, the International Commission on Illumination adopted 15 standard sky distributions which cover the whole spectrum of usual skies found in the world. Sky conditions of the same category would have similar sky luminance patterns and the corresponding climatic variables and indices would be within certain ranges. These analyses can help the identification of sky patterns. Once the standard skies are identified, the solar irradiance and outdoor illuminance at any surfaces of interest can be obtained for subsequent investigations and complicated expressions for inclined surface models are not required. Nevertheless, sky luminance data are not always readily obtainable in many parts of the world. The objective of this study is to identify the 15 CIE standard skies using various techniques based on Hong Kong measured data.

Sky luminance data and other climatic parameters collected by the measuring station in City University of Hong Kong from 1999 to 2005 were used to evaluate the CIE standard sky models. A set of 15 CIE standard skies which represent the general sky conditions was identified by using statistical approach. Based on these findings, a subset of 6 standard skies including overcast, partly cloudy and clear conditions were selected to represent the prevailing sky standards in Hong Kong.

In this study, sky conditions were identified using various techniques. Climatic parameters including solar altitude and ratio of zenith luminance to diffuse illuminance were selected for the characterization of sky conditions. Their characteristics, strengths and limitations in sky categorization were analyzed. A series of range of climatic

parameters was proposed to recognize the 15 CIE standard skies. The results were further compared with techniques proposed by other researchers. The proposed approach produced an overall RMSE of less than 33% without data reduction.

A special type of artificial neural networks, namely probabilistic neural networks was introduced. To investigate the feasibility of using probabilistic neural networks in sky type recognition, two different neural network models with the same test set were built to classify the 3 general sky conditions (i.e. overcast, partly cloudy and clear conditions) and the 15 CIE sky standards. Parametric analysis has also been carried out to investigate the essentiality of various climatic variables. The findings suggested that the neural network is an appropriate tool for sky classification and the ratio of zenith luminance to diffuse illuminance is the most essential input parameter to discriminate each of the standard sky type.

ACKNOWLEDGEMENTS

I am truly grateful to my thesis supervisor, Dr. Danny H.W. Li, for his supervision and inspiration to me on this research project. His indispensable suggestions and comments have made this thesis possible. I would like to express my sincerest gratefulness to my qualify panel members – Dr. Joseph C. Lam and Dr. Richard K.K. Yuen for their considerable support during my study.

Many thanks are due to the Head of Department, Prof. S. Kitipornchai, for his permission to carry out the study in the Department of Building and Construction, City University of Hong Kong. Also, I was supported by studentships offered by City University of Hong Kong. I am also indebted to all members of the Building Energy Research Group for their spiritual support and consideration. Special appreciations are given to Dr. Eric W.M. Lee who offered insightful advice and expertise on artificial neural networks, to Dr. Chris C.S. Lau, Mr. Ernest K.W. Tsang, Mr. Gary H.W. Cheung and Mr. Tony N.T. Lam who taught me a lot on day lighting and solar radiation.

Meanwhile, I would like to dedicate this thesis to the Management Team, Jockey Club Academy Hall, especially Mrs. Peggy B.W. Wong Chick, Mr. H.F. Kum, Mr. Rex T.W. Kwok, Mr. Mimo Y. Mao, Miss Ellen S.K. Wong. Without their endless encouragement and care, I could not have accomplished my study smoothly. Sincere thanks are expressed to my significant half, Miss Rita S.T. Lai for her tolerance and patience. Last but not least, I wish to extend the gratitude to my dear parents and sister for their silent concern and unconditional love.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ANN	Artificial neural network
BP	Back-propagation
CIE	International Commission on Illumination
DST	Daylight saving time
FA	Fuzzy adaptive resonance theory
FAM	Fuzzy ARTMAP neural network
GA	Genetic algorithm
GRNN	General regression neural network
HKO	Hong Kong observatory
HVAC	Heating, ventilating and air-conditioning
IDMP	International daylighting measurement program
NR	Normalization ratio
PCA	Principle component analysis
PDF	Probability density function
PNN	Probabilistic neural network
RMSE	Root mean square error
SSLD	Standard sky luminance distribution

NOMENCLATURE

α_s	solar altitude (rad)
β	azimuth of a sky patch (rad)
Γ	day angle (dimensionless)
δ	solar declination angel (rad)
θ	altitude of a sky element (rad)
σ	smoothing factor (dimensionless)
χ	scattering angle (rad)
a_v	luminous extinction coefficient (dimensionless)
c	cost of misclassification (dimensionless)
CLD	cloud cover (oktas)
D_v	horizontal diffuse illuminance (lux)
E_{SC}	solar constant (133.8klux)
E_v	horizontal extraterrestrial illuminance (lux)
G_v	horizontal global illuminance (lux)
H	hour angle (rad)
h	prior probability of the sample (dimensionless)
I_{nB}	direct-normal solar irradiance (Wm^{-2})
K	diffuse fraction (dimensionless)
K_t	clearness index (dimensionless)
L	latitude of the measurement location (rad)
L_γ	sky luminance of a sky element (cdm^{-2})

L_{mea}	measured sky luminance (cdm^{-2})
l_{pred}	predicted sky point luminance in relative form (dimensionless)
L_{pred}	predicted sky luminance (cdm^{-2})
L_Z	sky luminance at the zenith (cdm^{-2})
m_i	optical air mass (dimensionless)
N	number of readings (dimensionless)
n	total number of training pattern (dimensionless)
p	dimension of input vectors (dimensionless)
P_v	horizontal direct illuminance (lux)
r	eccentricity correction factor (dimensionless)
SH	sunshine hour (dimensionless)
T_{LK}	linke turbidity (dimensionless)
T_v	luminous turbidity (dimensionless)
$\mathbf{w}_i$	weight vector of the ith pattern neuron (dimensionless)
$\mathbf{x}_{Aj}$	j^{th} training vector from class A (dimensionless)
Z	zenith angle of a sky element (rad)
Z_S	zenith angle of the sun (rad)

Chapter 1

INTRODUCTION

Daylighting is an effective sustainable development strategy to alleviating the problems in energy, climate and environment, and improving the qualities for visual comfort and health [Kittler et al., 1999; Kambezidis et al., 1998]. It has been recognized as an important design element in buildings throughout history. Proper daylighting design reduces the energy demand and creates dynamic interior space supportive of human performance and health. On the contrary, improper system design causes visual discomfort, depression and a waste of energy [Leslie, 2003]. Energy efficient technique on lighting is one of the important criteria to access the performance of buildings. Nevertheless, the study of daylighting design has fluctuated throughout history [Reinhart and Selkowitz, 2006].

For over half a century, daylighting was treated as a primary light source of buildings. Since the installation cost of the daylighting devices such as laser cut panels and light pipes is normally higher than the fluorescent lighting systems, the investigations of daylighting designs were ceased and thus the indoor illuminance in most of the

buildings depends on the electric energy. Although there was a growing awareness for natural light in the early 1970s due to the energy crises and rapid increases in energy prices, the trend could not be maintained in the next several decades. The impacts of global warming and climate change are emerging, nowadays, integrating daylight with artificial light has become a major topic again in building system designs. It is believed that daylight is superior to fluorescent light in terms of energy and environment, human performance, visual comfort and health.

Energy and Environment

In subtropical Hong Kong, Lam et al. [2003] reported that the major electricity consuming component in fully air-conditioned commercial buildings is heating, ventilating and air-conditioning (HVAC) systems, representing about 30-60% of total annual electricity use. The electric lighting systems is the second, accounting for 20-30%. For commercial sector in other developed countries, the electricity consumption for the artificial lighting in USA and Canada is about 35 and 30% respectively [Galasiu and Veitch, 2006]. Hence, the rational utilization of daylight has a great potential in energy saving.

Recently, there is a tendency in integrating daylight with electric light for saving energy in buildings [Jenkins and Newborough, 2007]. Energy can be saved by either automatically dimming down or manually turning off the artificial lights whenever daylight is sufficient to provide background illuminance. Also, it is particularly attractive in hot climate since proper daylighting designs can reduce electricity bills paid for HVAC systems due to less heat dissipation from light fittings [Zmeureanu and Peragine, 1999; Li et al., 2006b].

Based on the results of simulation programs and field measurements, several researchers reported that about 30 to 77% of lighting and cooling energy can be saved by using different daylighting control schemes [Doulos et al. 2008; Ihm et al., 2008; Onaygil and Guler, 2003]. For Hong Kong, Li et al. [2006b] reported that the annual energy savings in electric lights for the open plan office under the proposed daylighting strategy was 365kWh, representing a 33% reduction in energy use for electric lighting.

Apart from the energy savings in both lighting and cooling, natural light is able to preserve the environment. Over the past few decades, there has been a growing concern about climate change and global warming. In most developed countries, energy is generated by burning fossil fuels and the process simultaneously emits air pollutants and greenhouse gases such as sulphur dioxide, carbon dioxide and nitrogen oxide. Irreversible damages on environment have been caused by those air pollutants which still keep on increasing [Bodart and De Herde, 2002]. Promoting daylight is one of the feasible solutions to reduce the fossil fuels consumption as well as the greenhouse gases emission.

Human Performance

Daylight has a significant psychological effect on human being. Our daily life and activities are still related to the daylight even though electric lights are available. Normally, the human rhythm of work and rest do follow the path of the sun [Kobav and Bizjak, 2003]. People desire good natural lighting in their living and working environments. Daylighting has been recognized to improve the productivity of the human [Smiley, 1996]. A number of studies reported that daylighting in schools can not only significantly increase the academic performance of students, but also promote

better health and physical development [Plympton et al., 2000]. Heschong [2002] listed the possible associations between daylight and improved performance of students such as the improved visibility due to higher illuminance levels and better light quality. Leslie [2003] claimed that the improved performance may attribute to the suppression of melatonin which is affected by exposure to the daylight. To maximize the use of daylight and increase the human productivity, meanwhile, daylight saving time has been introduced to shift the human activity patterns [Aries and Newsham, 2008].

Visual Comfort and Health

Daylight is the only light source that is most closely matches human visual responses. It provides a good visual environment which affects one's visual ability to see an object inside a room. It has been proved that human visual responses are directly related to the daylighting designs in buildings [Lam, 1985]. In general, summer has longer day length and shorter nighttime, with a reverse trend in winter. A number of researchers stated that most of the human are emotionally susceptible to seasonal variation in daylight, and up to 9.7% of people may suffer from seasonal affective disorder [Magnusson, 2000; Kasper, 1989] which is a clinical subtype of major depressive disorder in which the onset and remission of depressive episodes recur at certain times of the years [Lundt, 2004]. The sufferers should be treated by light therapy to offset the depressive reactions. The influences of daylight on the human were described in detail by Webb [2006] who concluded that daylight controls the circadian rhythm of hormone secretions and the body temperature cycle with implications for sleep/wake states, alertness, mood and behavior.

1.1 Background

The determination of the interior daylight levels is a key on daylighting designs and planning of the illumination systems [Aoki et al., 1997]. The first step towards designing a building to utilize daylight for illuminating its interior is to acquire information on the amount of daylight available outside the building. An accurate estimation is to acquire not just the total amount of external illuminance levels available, but also the distributions of luminance across the sky vault [Li et al., 2006a]. The sky luminance distributions are affected by various factors such as solar position, atmospheric turbidity and the air pollution as well as cloud amount, type and pattern. These can affect unpredictably sunlight and skylight [Kittler et al., 1997b]. Over the past few decades, several empirical models have been established to determine the sky luminance distributions under various sky conditions [Moon and Specer, 1942; Perez et al., 1993; Littlefair, 1994]. However, only two of the sky conditions, CIE standard overcast sky and CIE standard clear sky, were considered in most of the daylight simulation programs and modeling. Kobav and Bizjak [2003] pointed out that it is no longer satisfactory to determine the indoor daylight illuminance level with only two sky conditions. To have a better result in daylight calculations, therefore, sky luminance distributions between the two CIE standard skies should be standardized properly and systematically.

In 2003, International Commission on Illumination (CIE) adopted 15 standard skies [CIE, 2003] which cover the whole probable spectrum of skies found in nature. The set includes 5 overcast, 5 partly cloudy and 5 clear skies. Numerous researchers reported that the 15 CIE standard skies provide a good overall framework for representing the

actual sky conditions [Tregenza, 1999; Li et al., 2003; Ng et al., 2007]. To identify the set of standard skies, the luminance distributions of individual standard skies are modeled and compared with the scanned sky luminance readings. Once the skies are interpreted, the basic outdoor illuminance at various inclined surfaces can be obtained for subsequent investigations. Nevertheless, luminance distributions for the whole sky vault are far from being widely available [Li et al., 2004b; Chirarattananon and Chaiwiwatworakul, 2007]. Even in some developed countries, basic daylight measurements particularly sky luminance distributions are only available in several cosmopolitan cities.

In interpreting sky conditions, climatic parameters are the critical weighting factors to indicate the degree of sky clearness such that sky distributions can then be classified. Sky conditions of the same category would have similar sky luminance distributions and the corresponding climatic parameters and indices would be within certain ranges [Li et al., 2004a]. Such analyses can help the identification of sky patterns while complicated mathematical expressions to model the sky distributions are not required. Research works on sky classification using meteorological parameters can be found in many parts of the world [Barbaro et al., 1981; Perez et al., 1990; Long and Ackerman, 2000; Li and Lam, 2001; Li et al., 2004a; Li et al., 2004b; Li and Lau, 2007; Younes and Muneer, 2007]. All the works conducted by the researchers have taken a great step forward in sky identification. However, these studies were mainly focused on categorizing the 3 general sky conditions, namely overcast, partly cloudy and clear sky. Specific investigations on classifying the 15 CIE standard skies are rare in literature. Kittler et al. [1998] suggested that the ratio of the zenith luminance to the horizontal diffuse illuminance (L_z/D_v) can be used to discriminate the daylight measurement into one of

the 15 standard sky patterns. Later, Bartzokas et al. [2003; 2005] applied this climatic index to study the daylight climate in Athens and Bratislava, but the performance of the approach was not further evaluated. Also, it has been pointed out that ambiguous results were caused when defining the sky types in high solar altitude (α_s).

Although there are a number of appropriate climatic parameters (both solar irradiance and outdoor illuminance) identifying the daylight climates [Kittler and Darula, 2002], only solar irradiance measurements such as clearness index (K_t) and diffuse fraction (K) were normally used to express the sky conditions. Therefore, study focus on daylight measurements for sky classification is essential. To achieve this, several approaches are suggested in this study to provide simple and precise tools for sky classification.

1.2 Aims and Objectives

The primary goal of this research study is to classify the 15 CIE standard skies using synoptic meteorological parameters and artificial neural networks (ANNs). To achieve this goal, the following specific objectives are addressed:

i. To analyze the characteristics and variations of the proposed climatic parameters under 3 typical sky conditions.
ii. To define the 15 CIE standard skies and a subset of prevailing standard skies in Hong Kong by using statistical approach.
iii. To develop and evaluate the sky classification tools for recognizing the 15 CIE standard skies using common climatic parameters and ANNs.

iv. To compare the performance of various sky classification approaches against scanned sky luminance data measured in Hong Kong.

v. To investigate the essentiality of daylight variables for sky classification using ANNs.

1.3 Thesis Outline

This thesis is subdivided into six chapters, each covering the major step to develop the proposed approaches for sky classification. The works of each chapter are listed as follows:

Chapter 1, Introduction –

To discuss the impacts of daylighting designs against building energy, human performance as well as visual comfort and health. The aims and background of this study are described. A brief overview of each chapter is introduced.

Chapter 2, Daylight Measurements and General Results Analysis –

To provide some background information about the daylight climates in Hong Kong. The measuring station constructed at the City University of Hong Kong is described. Data quality control tests and the general results of daylight measurements are the main subjects of this chapter.

Chapter 3, Climatic Parameters and CIE Standard Sky Model –

To suggest the climatic variables for the current study. Their strengths and impacts on

sky classification are discussed. Furthermore, the 15 CIE standard skies for Hong Kong are defined based on the best-fitting approach. Subsets of 6 prevailing sky standards as well as 3 sky standards for Hong Kong are selected and investigated via graphical representations and statistical methods.

Chapter 4, Sky Classification by Climatic Parameters –

To present the statistical approach on the classification of standard skies using various climatic parameters, the sky classification approach recommended by Bartzokas et al. [2003; 2005] is briefly introduced and evaluated against measured sky luminance data in Hong Kong. Then, a series of range of climatic parameters is suggested for recognizing each of the sky standards. The results and performances of the proposed approach are presented afterwards.

Chapter 5, Sky Classification by Artificial Neural Networks –

To compare the two types of ANNs, namely neural network with back-propagation (BP) learning algorithm and probabilistic neural networks (PNNs), which are commonly used in pattern classification problems. The performance of the PNNs with majority voting scheme for categorizing the 3 general sky conditions and 15 CIE standard skies are evaluated. Parameter analysis is performed to examine the importance of each input parameter of the PNNs for sky categorization. A comparative investigation between the climatic parameter and PNNs approaches are presented at the end of this chapter.

Chapter 6, Conclusions and Recommendations –

To summarize the major findings in this study. The limitations of the research and further studies on sky classification are illustrated.

Chapter 2

DAYLIGHT MEASUREMENTS AND GENERAL RESULTS ANALYSIS

Data measurement is the most effective and accurate approach of setting up the databases of daylight [Jeter and Balaras, 1990]. However, there is a lack of daylight illuminance measurements in many parts of the world [Kittler et al., 1992; Lin and Lu, 1992]. In order to set a guideline for the worldwide researchers and professional organizations to measure daylight availability parameters, CIE designated the year of 1991 as the "International Daylighting Measurement Year" which represents the first year of the "International Daylighting Measurement Program" (IDMP) [CIE, 1994]. This program not only provides a guideline for the daylight measurements and the instrumentation requirements, but also suggests various recommendations for data quality control, archiving and dissemination. Under the IDMP, there are two types of measuring stations:

i. The General Class Station –

   Basic daylight measurements should be recorded at the station including

horizontal global and diffuse for both illuminance and irradiance, vertical global illuminance on four cardinal surfaces facing the north, east, south and west, and sunshine hour.

ii. The Research Class Station –

Apart from the measurements recommended at the general class station, sky luminance distributions, total cloud amount, zenith luminance, dew point temperature, air temperature, vertical global irradiance facing the four cardinal directions should also be recorded.

Hong Kong Observatory (HKO) is a pioneer to record the meteorological measurements in Hong Kong since 1958 [Sham, 1964]. At the early stage, only daily global solar radiation on horizontal surface was taken. Since the solar radiation is an essential element to estimate the heating and cooling loads of buildings, and design of energy efficient systems, the measuring equipments were upgraded to record both of the daily and hourly global solar radiation in 1978 [Lau, 1989]. Nevertheless, no such measurements on outdoor illuminance are recorded until now.

To define the daylight climate in Hong Kong, a measuring station was established at the City University of Hong Kong in 1991. Initially, only measurements of solar irradiance and illuminance (global and diffuse) on a horizontal surface were made. In 1996, the measurements were extended to vertical global irradiance and illuminance on four cardinal surfaces facing north, east, south and west. The measuring station was upgraded in 1999 with the installation of a sky scanner to record the radiance and luminance distributions of the whole sky vault. In July 2006, four solar irradiance

sensors were mounted with an inclination angle of 45^{o} facing north, east, south and west. In October 2006, another eight sensors were installed with tilt angles of 30° and 60° facing the northeast, south east, southwest and northwest to record the global solar irradiances.

This chapter provides some background information about the daylight climate in Hong Kong. The measured 10-minute data during 1999-2005 are gathered to form a seven years database for further analyses. Data quality control and correction tests are performed. Moreover, the general measured results on daylight illuminance are analyzed based on the statistical techniques. Sky luminance distributions for the 3 typical sky conditions are examined at the end of this chapter.

2.1 Hong Kong Climate

Hong Kong has a total area of about 1,000km^2 and is situated along the southern coast of China within the subtropical region, with latitude and longitude of 22.3°N and 114.2°E, respectively. It has distinct seasonal changes in its weather due to its location on the south-east coast of the Asiatic continent opposite a vast expanse of the ocean. Winter months are between December and February. Mean temperature is around 15-18°C with the monsoon coming from the north and north-east. Spring is from March till early May, and is usually cloudy with periods of light rain. The summer season spans from late May till September. The monsoon wind blows from the south and south-east with an average temperature of 27-29°C. It is hot and humid with occasional showers and thunderstorms. At times, typhoons strike Hong Kong and bring heavy rain and

strong winds. Autumn is short and normally runs from October to November. Sunny bright skies dominate with dry conditions and a mean temperature of 25°C.

2.2 Measuring Station

The measuring station is located on the roof-top and a laboratory beneath the roof (top-floor level) of the City University of Hong Kong. All sensors were installed on the roof-top in a position relatively free from any external obstructions, and readily accessible for inspection, general cleaning and maintenance. Apart from the outdoor measuring station, there is a laboratory (shown in Figure 2.1) located at the top floor of the building. The electronic apparatus such as signal amplifier, solar integrators and computers were placed inside the laboratory. The data collection started just before sunrise (6:00a.m.) and ended after sunset (6:00p.m.) each day. All measurements were recorded in true solar time (TST), instead of local time. This facilitates the computations involving α_s for the extraterrestrial irradiance/illuminance on the unit horizontal surface and the subsequent comparison of the data at other locations.

Figure 2.2 presents an overview of the apparatus set-up in the measuring station. As mentioned, a total of eighteen pyranometers, six illuminance sensors and a sky scanner were installed for daylight measurements. All the data signals were transferred from the sensors to six computers located inside the laboratory. For the solar irradiance measurements, six pyranometers which record the horizontal global and diffuse irradiances as well as the four vertical irradiances were averagely split into three groups and each group was linked with a solar integrator and computer. The remaining twelve

pyranometers were connected to a computer (PC6) with software, LabVIEW version 7.1, via the signal amplifier and data acquisition module. For the outdoor illuminance measurements, the six illuminance sensors which record the horizontal global and diffuse illuminance as well as the four vertical illuminance sensors were simply linked in series to a computer (PC4). Apart from the irradiance and illuminance measurements, the sky radiance/luminance distributions measured by the sky scanner was stored in another computer (PC5). In this study, only measurements of outdoor illuminance and sky luminance distributions were used for analysis.

Figure 2.1 Measuring equipments inside the laboratory

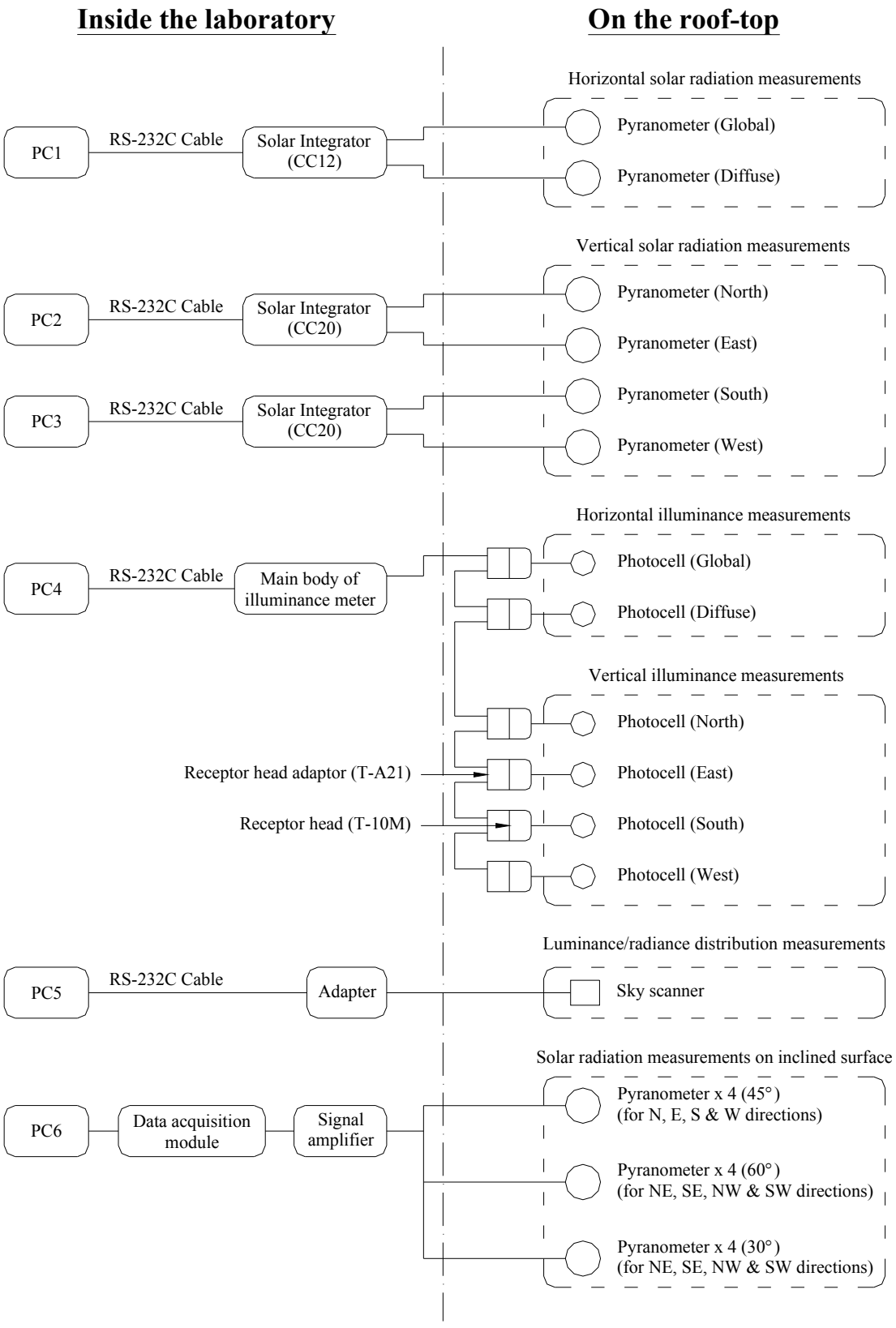

Figure 2.2 Block diagram of the apparatus set-up in the measuring station

2.2.1 Outdoor daylight illuminance

Before introducing the measuring equipments, it is important to define the component of outdoor illuminance first. The horizontal global illuminance (G_v), which consists of diffuse and direct components, is the total amount of outdoor illuminance falling on a horizontal surface. The illuminance arrives after being scattered either in the atmosphere or off adjacent surfaces at ground level is regarded as horizontal diffuse illuminance (D_v). By subtracting D_v from G_v, horizontal direct illuminance (P_v) is obtained, representing the illuminance of the direct component arrives directly from the sun after passing through the unobstructed atmosphere.

The measurements of outdoor global and diffuse illuminance were made by means of illuminance sensors (T-10M) manufactured and calibrated by Minolta of Japan. The silicon photocells with cosine and colour corrections measure illuminance up to about 300klux with an accuracy of ±2%. A multi-point illuminance measurement system was used. The diffuse illuminance sensor was fitted with a shadow ring (CM121) 55mm wide and 620mm in diameter, which shades the thermopile reflections. The shadow rings were painted black to minimize the effect of multiple reflections. According to the instruction manual published by Kipp & Zonen [Kipp & Zonen], regular adjustment of the shadow ring position should be done every two days. However, a longer time interval between adjustments is possible in most seasons. A data-measurement package was used to capture the measured readings from the main body adapter and the data were then fed into a computer for storage. All the outdoor illuminance measurements were being collected every minute. Figure 2.3 to Figure 2.5 show the illuminance sensors used for the corresponding outdoor illuminance measurements.

2.2.2 Sky luminance

The sky luminance distributions were recorded by means of a sky scanner (EKO MS 300LR) manufactured and calibrated by the Japanese company EKO. The photo of sky scanner is displayed in Figure 2.6. The sensor head rotates in altitude and azimuth to measure the luminance at 145 circular sky patches by scanning the sky hemisphere. The full view angle of the scanner is 11º that allows each sky patch to be treated as a point source with negligible error [Tregenza, 1987; Tregenza and Sharples, 1995]. The measurement method excludes certain regions of the sky vault and gives a sky coverage of approximately 68%. The sequence of the 145 measurement points for sky scanning is shown in Figure 2.7. The sky grid pattern illustrated in Figure 2.7 (not the measurement pattern) was suggested by Tregenza [1987] such that the whole sky-dome can be considered for subsequent analysis. Later, CIE recommended such subdivision for researchers to design and build a scanning luminance meter [CIE, 1994].

The important parts of the sky scanner were housed in a weatherproof casing allowing continuous outdoor operation. The sky radiance/luminance data collected from the scanner were recorded on a computer placed inside the laboratory on the top floor. A visual basic computer program was used to capture and transmit the measured data. To safeguard the sensor, the scanner did not record luminance data greater than 35kcdm^{-2} ($\approx$ 210Wm^{-2}sr^{-1}) by using an automatic shutter [EKO, 1998]. Each scanning period was about 4 minutes and the measurements were taken every 10 minutes.

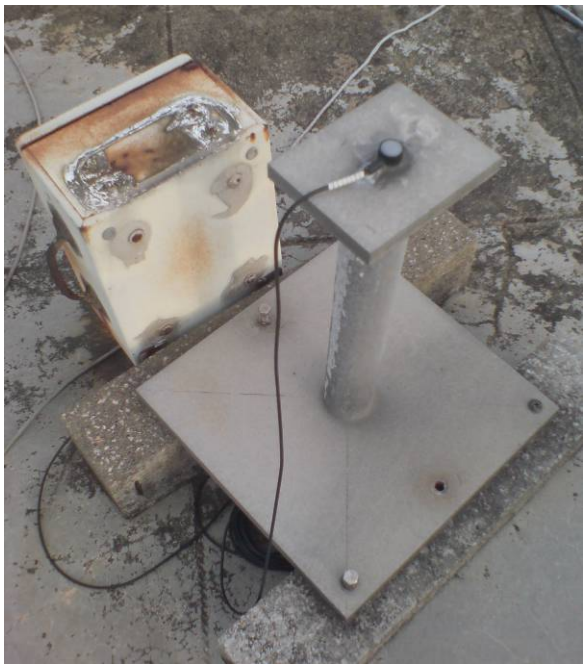

Figure 2.3 Illuminance sensor used for horizontal global illuminance measurement

Figure 2.4 Illuminance sensor fitted with shadow ring used for horizontal diffuse illuminance measurement

Figure 2.5 Illuminance sensors used for vertical global illuminance measurement

Figure 2.6 Sky scanner used for sky luminance distribution measurement

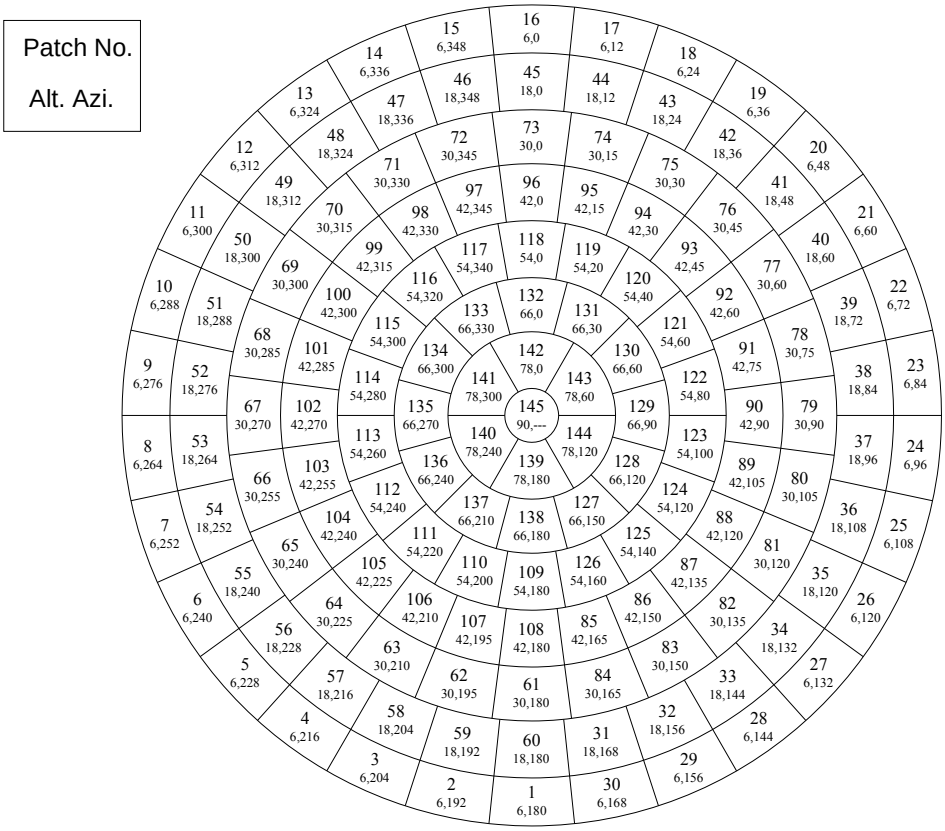

Figure 2.7 The sequence of the 145 measurement points for sky scanning (lower numbers left to right: the altitude and azimuth angles of the centre of sky points)

There are several causes affecting the accuracy of the measured sky luminance data. Firstly, the division of the sky into 145 circular angular patches can avoid any double counting, but it excludes certain regions of sky vault. Secondly, the measured data were based on discrete results rather than continuous analytical functions. The real sky luminance patterns are influenced by various circumstances such as changing solar positions, turbidity and pollution content of the atmosphere as well as the distribution of clouds on the sky dome that cause various interactions of the sunlight and skylight. Since the sky distributions are changing from time to time, sky luminance between adjoining measurement points may vary significantly. Thirdly, substantial variations in

sky luminance may occur within each record. Besides, for "out of range" measurements (points close to the solar position under non-overcast skies), estimation of the sky luminance was made from a simple average of the luminance at nearby points; such a conversion, however, could introduce data distortion [Li et al., 2005].

2.3 Data Corrections

Undoubtedly, data corrections are the key steps prior to data analysis. It is important to check all recorded data completely against inconsistencies, otherwise, the results of the study will be affected accordingly [Hay, 1993]. For both general class and research class IDMP station, CIE [1994] provided some useful guidelines and recommendations to perform data corrections effectively.

2.3.1 Data quality control

The purposes of data quality control are not only eliminate spurious data and erroneous measurements, but also detect the instrument problems. Common errors during measurements included shadow ring misalignment, cosine effect of the sensor, large calibration drifts and instrument malfunctions [Li et al., 2005, Younes et al., 2005]. In this study, the following quality control tests were adopted based on CIE Guide to Daylight Measurement [CIE, 1994].

i. Rejecting all data with α_s less than 5º.

ii. Rejecting all data with horizontal global irradiance less than 20 Wm^{-2}.

iii. Rejecting all global data that was greater than the corresponding extraterrestrial

solar components.

iv. Rejecting all diffuse data that was greater than the corresponding global values.

v. Rejecting all diffuse data that was greater than half of the corresponding extraterrestrial solar components (because of the improper shadow-ring adjustment).

vi. Rejecting all data when the direct-normal values [i.e. (global - diffuse)/sin α_s] exceeded the corresponding extraterrestrial solar component.

vii. Rejecting all horizontal and vertical data when the computed vertical direct beam data was greater than the corresponding measured vertical global component.

viii. Rejecting all sky luminance data when the differences between the corrected D_v and the corresponding integrated D_v based on the 145 scanned sky luminance points were greater than 30%.

2.3.2 Diffuse illuminance correction

Diffuse illuminance was measured using shadow-ring to eliminate direct sunlight. Occasionally, the shadow-ring may be misplaced, resulting in diffuse readings being closed to that of global under non-overcast skies. Moreover, the shadow-ring does block off a significant portion of the sky-diffuse illuminance and thus some corrections must be applied to obtain the true value. Vartiainen [2000] indicated that proper corrections of the diffuse daylight measurements are an important step for model analysis. The corrections were usually made by assuming that the sky is isotropic [Drummond, 1956; LeBaron et al., 1980; Painter, 1981]. However, because of the anisotropy of diffuse sky, with its maximum being close to the sun, the isotropy-sky-approach can lead to errors. As a result, a number of models were developed or modified to determine the correction

factor of the anisotropic sky diffuse components [Kudish and Ianetz, 1993; LeBaron et al., 1990; Littlefair, 1989]. To create a reliable database, the diffuse component measurements obtained with the shadow-ring were corrected using the model described by LeBaron et al. [1990].

2.4 General Data Analysis

After passing through the quality control tests and the diffuse illuminance corrections, a reliable quality controlled database was obtained. In total, there were about 68,000 set of 10-minute data. Approximately 118,000 of measurement data were rejected, representing about 63% of the total set of 186,000 data. Similar rejection rate was obtained by Ng et al. [2007]. The high rejection rate may be attributed to the equipment errors and instrument malfunctions. They were based on the measurements during the seven years period from 1999 to 2005 at City University of Hong Kong. The following items were included in the database.

(a) Measurement date and time

(b) Solar altitude and azimuth

(c) Horizontal global illuminance

(d) Horizontal diffuse illuminance

(e) Horizontal direct illuminance

(f) Vertical global illuminance on planes facing north, east, south and west

(g) Sky luminance distribution over the sky vault

(h) Zenith luminance

In the following sections, the measured outdoor illuminance and scanned sky luminance data were gathered to analyze the daylight climate of Hong Kong. The frequency distributions and seasonal variations of the outdoor illuminance were investigated using statistical methods. Moreover, the general characteristics of the 3 typical sky conditions (i.e. overcast, partly cloudy and clear sky) were discussed. In order to have a better view of the sky patterns, the measured sky luminance under the 3 typical sky conditions were displayed graphically.

2.4.1 Horizontal outdoor illuminance

The frequency of occurrence for the 10-minute G_v and D_v at an interval of 1klux was computed and showed in Figure 2.8 and Figure 2.9, respectively. It should be noted that the high occurrences at low G_v are attributed to the overcast sky and conditions near sunrise and sunset hours. The frequency distribution decreases gradually from low to high G_v of 129klux. A general inspection of Figure 2.9 revealed that D_v between 4klux and 26klux are quite evenly distributed. Beyond 26klux, however, a decreasing trend is observed. The distribution drops gradually to 65klux, about half of the peak value of G_v.

In most applications, the frequency distributions provide a useful reference for researchers and professional engineers to study and design the artificial lighting of the buildings. Figure 2.10 presents the cumulative frequency distributions for both G_v and D_v in Hong Kong. It can be seen that G_v exceeds 25klux for 55% of the time. That means daylighting alone can provide a room with 2% of daylight factor and an indoor design illuminance of 500lux for 55% of time no matter what the sky condition is. However, only 39% of time that the room is able to maintain the same daylight factor and indoor design illuminance as the direct sunlight is not available.

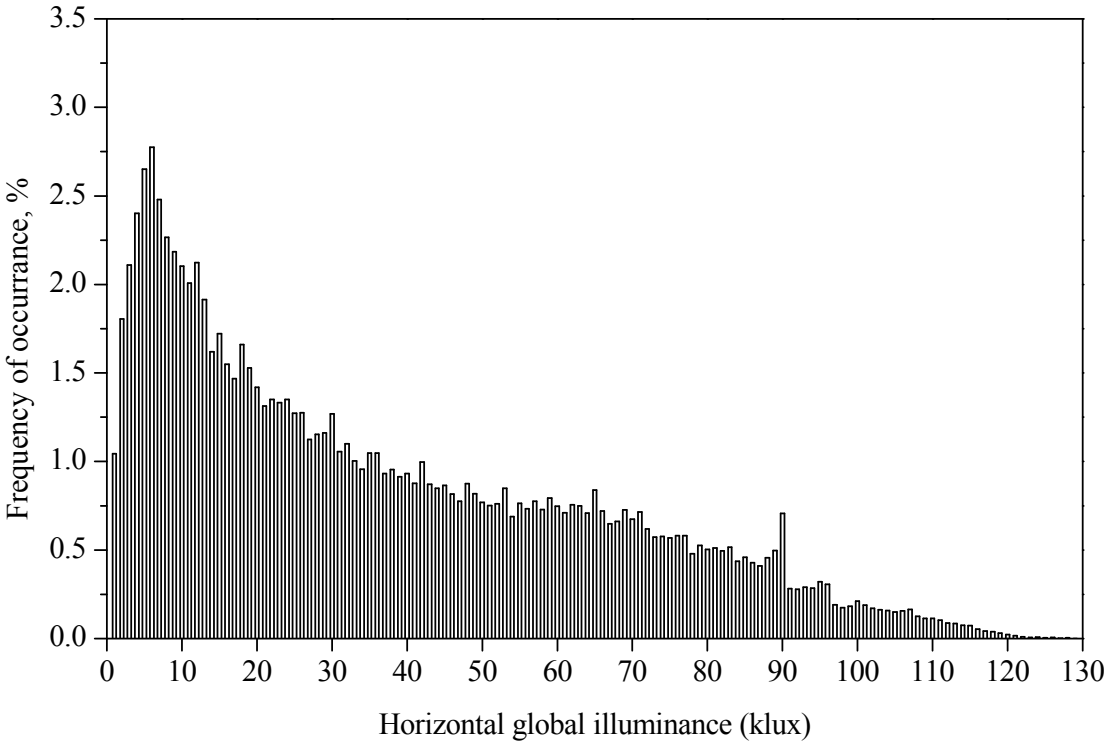

Figure 2.8 Frequency of occurrence of horizontal global illuminance

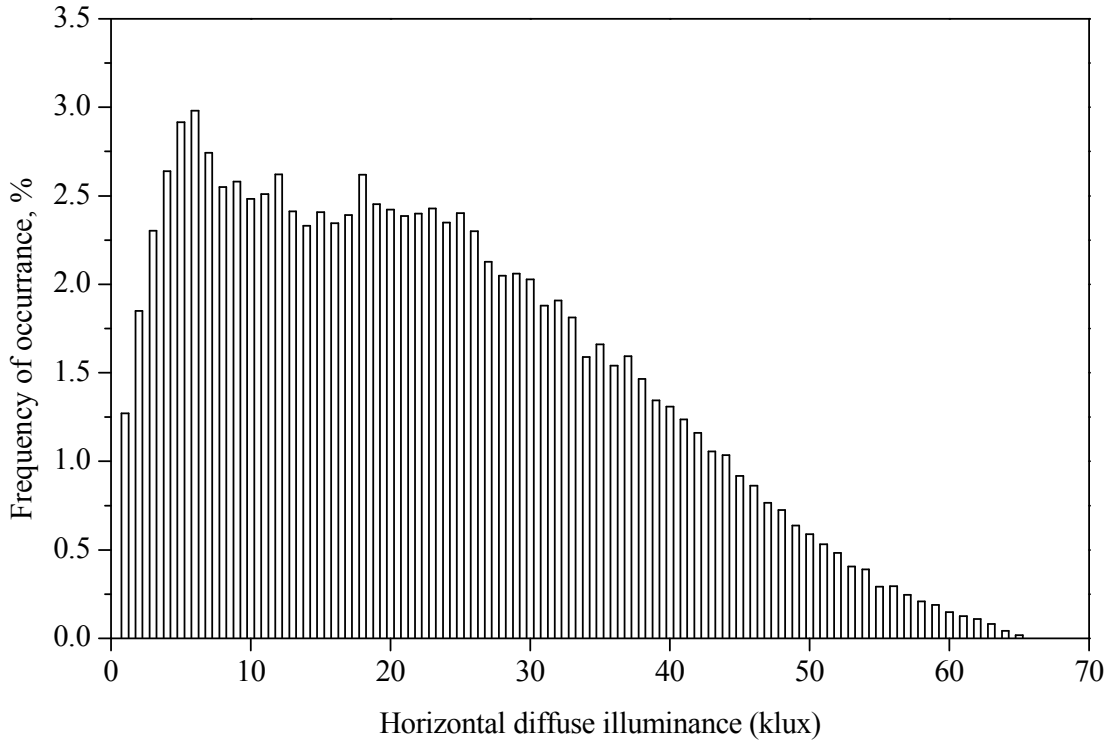

Figure 2.9 Frequency of occurrence of horizontal diffuse illuminance

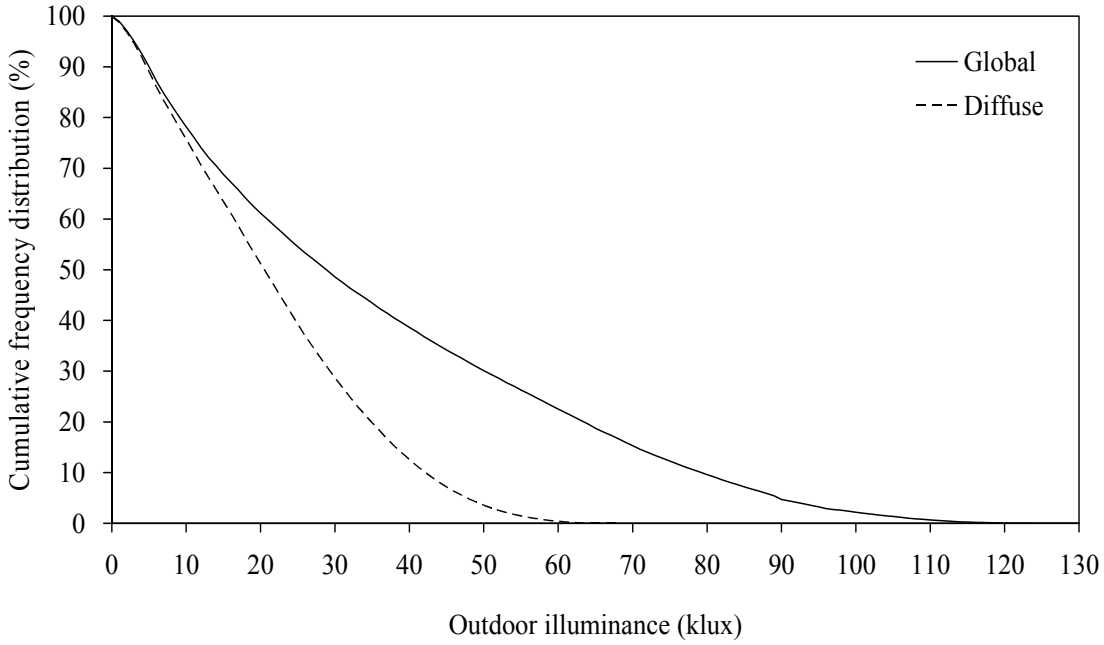

Figure 2.10 Cumulative frequency distribution of horizontal global and diffuse illuminance

Figure 2.11 and Figure 2.12 present both the monthly-average-hourly G_v and D_v during 1999-2005 respectively. As expected, the outdoor illuminance on horizontal surface obtains the greatest values at solar noon because of the highest α_s. G_v at solar noon are from 49klux in January to just over 66klux in July. The maximum and minimum D_v around solar noon is 37klux in May and 26klux in December respectively. A detailed examination of the two isopleth diagrams revealed that large amount of P_v occurs at solar noon in the months of July, indicating clear sky conditions in summer. The low P_v during the midday in March and April indicates the overcast days in spring. In general, both the G_v and D_v are symmetrical with respect to solar noon.

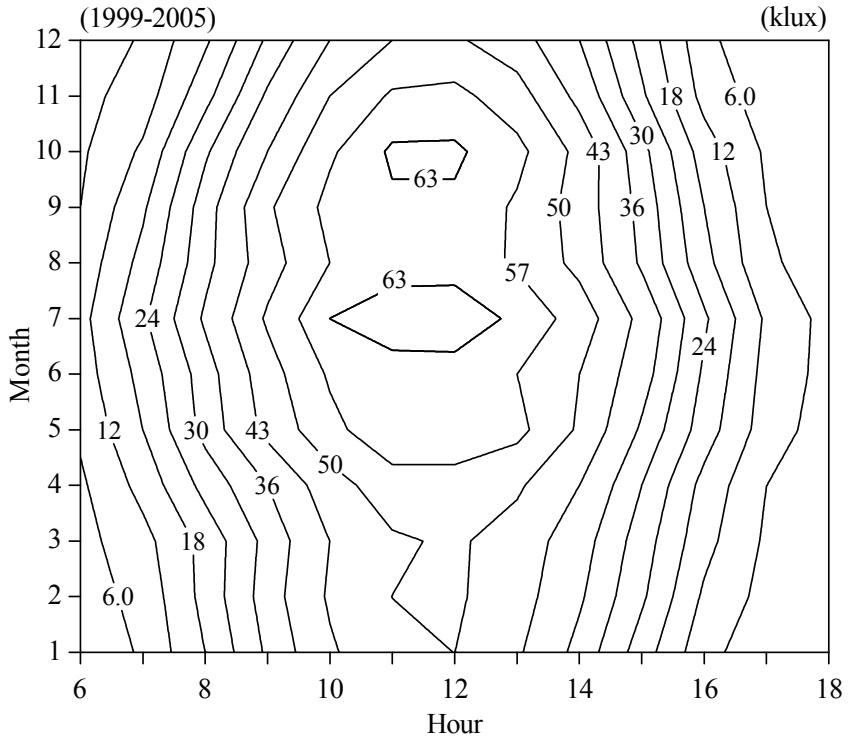

Figure 2.11 Isopleth diagram of global outdoor illuminance

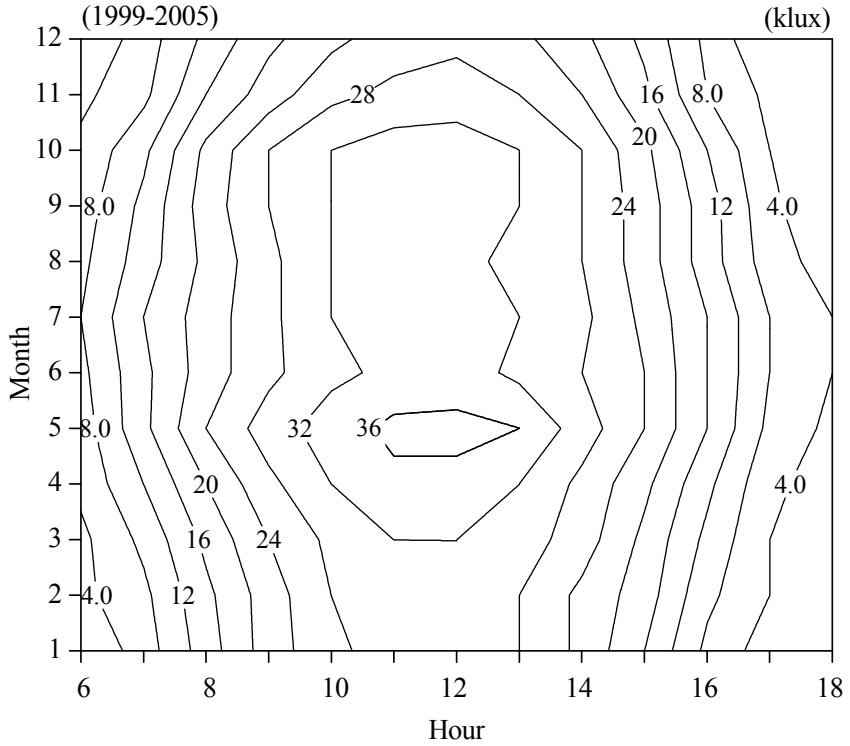

Figure 2.12 Isopleth diagram of diffuse outdoor illuminance

2.4.2 Vertical outdoor illuminance

To further examine the daylight climate of Hong Kong, the seasonal variations of the vertical outdoor illuminance were investigated. Monthly-average-hourly outdoor illuminance for the four vertical planes (north, east, south and west) for January, April, July and October of 1999 to 2005, representing winter, spring, summer and autumn, respectively, are shown in Figure 2.13. Large seasonal variations and irregular patterns are observed in those figures. The main cause of daily variations of vertical illuminance is the change of solar position. In the absence of atmospheric absorption, the outdoor illuminance on a vertical surface facing the sun would be the highest when the sun is near the sunrise and sunset hour. For instance, the maximum outdoor illuminance always occurs at the east-facing plan in the early morning, no matter what the season is.

For the south-facing plane, due to the sun-path of Hong Kong (shown in Appendix A), the surface receives outdoor illuminance from mainly direct component in winter to mainly diffuse component in summer, resulting in a high illuminance level in January (47klux around solar noon) and low illuminance level in July (22klux around solar noon). Likewise, the lowest outdoor illuminance values in north-facing surface could be attributed to the lack of the direct components. The outdoor illuminance for east and west-facing planes are quite symmetrical with respect to solar noon throughout the year. The peak outdoor illuminance for north, east, south and west-facing planes are about 18klux, 36klux, 49klux and 41klux, respectively.

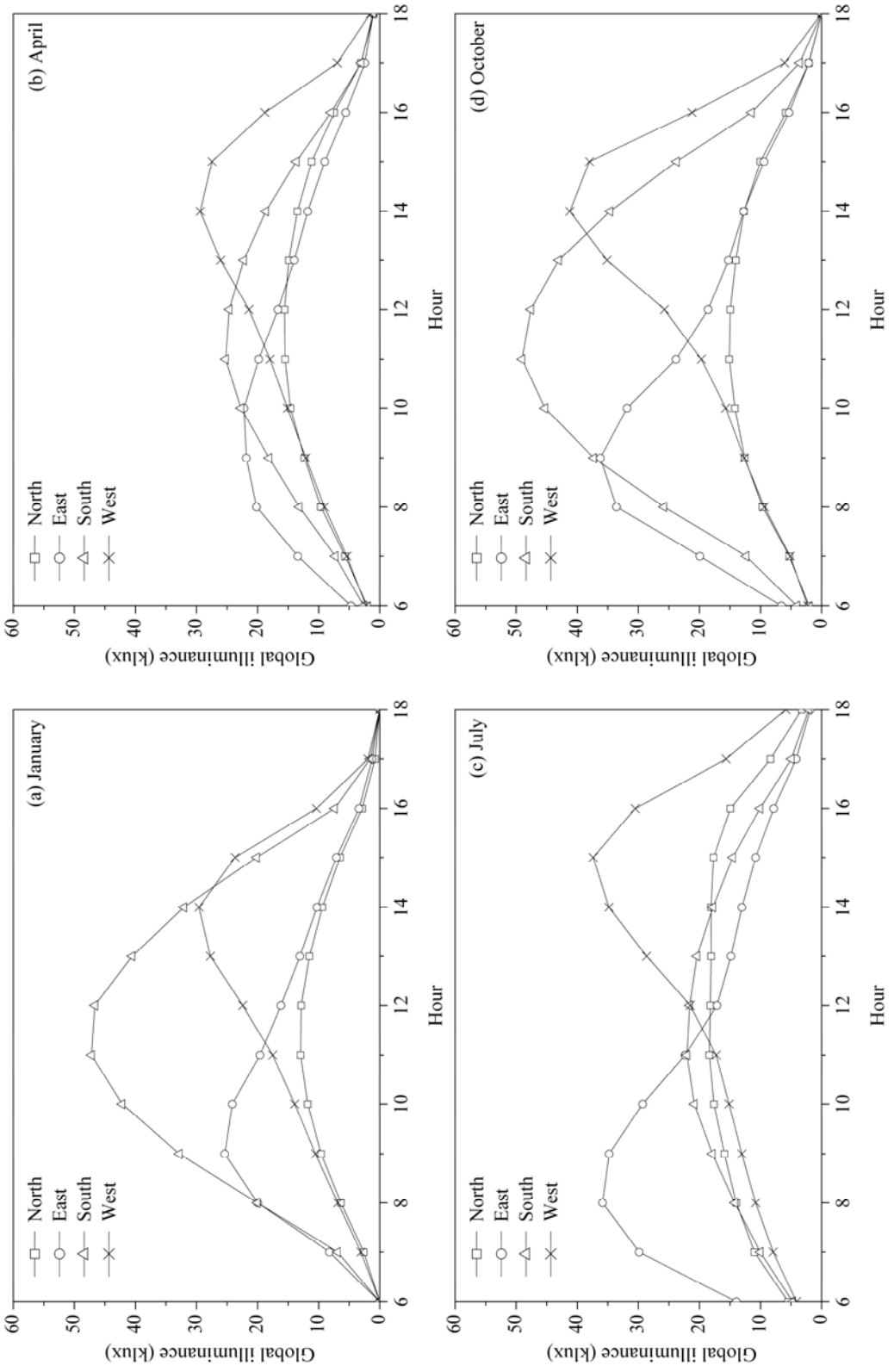

Figure 2.13 Monthly-average-hourly outdoor illuminance on the vertical surfaces for January, April, July and October

2.4.3 Sky luminance

Sky luminous distributions depend strongly on the weather conditions and the local climate, and it varies with the solar position from time to time [CIE, 2003]. As a result, the analysis of sky luminance distributions is complicated. Nonetheless, sky conditions of the same category would have similar sky luminance distributions and the corresponding climatic parameters and indices would be within certain ranges. Li and Lau [2007] reported that sky conditions can be categorized into overcast, partly cloudy and clear types by using L_z/D_v and G_v/E_v. In this section, three set of data were selected to represent the luminance distributions under the three typical sky conditions. Graphical representation is the best method to present the full scan measurements of the sky vault because of the better visualization of the actual distribution [Kittler et al., 1992].

Overcast sky

The standard overcast sky is often considered to provide the worst conditions for daylighting design and performance analysis [Longmore, 1975; Li et al., 2001]. The sky totally covered with a layer of clouds as well as the vision of the sun is completely impeded can be regarded as overcast sky [Enarun and Littlefair, 1995]. Under the CIE standard overcast sky proposed by Moon and Spencer [1942], the sky luminance distribution is not only symmetrical about the zenith, but also independent on the solar position. Figure 2.14 depicts the sky luminance distribution for an overcast sky. There is clearly no disturbance of the luminance pattern around the solar position. A remarkable peak sky luminance occurs at the zenith of around 3.6kcdm^{-2} while the lowest sky luminance appears at the horizon, ranging from 0.3 to 1.3kcdm^{-2}. This large sky luminance variation at the horizon could be attributed to the different reflectivity values

from the surroundings at low elevations. The sky luminance differences at different orientations of the same elevation become small as the elevation increases.

Clear sky

Investigating the sky luminance on clear days is also important for estimating the peak cooling load and sizing the air-conditioning equipment capacity in buildings, particularly for those situated in places where clear days occur more frequently. The luminance distribution of the clear sky depends on the solar position and the scattering of light in the atmosphere [Littlefair, 1994]. The peak sky luminance generally appears in the direction of the sun, decreases rapidly with the distance from the sun increases. Figure 2.15 shows the sky luminance distributions for a clear sky. As mentioned, the sky luminance near to the sun can be several hundred times of the other sky elements, and thus the sky luminance level is limited to 35kcdm^{-2} for better visualization. As expected, the maximum sky luminance appears at the south orientation that is the solar position. The smallest luminance of 1.8kcdm^{-2} is found at the horizon for the north. For the same elevation, the maximum sky luminance of 12.2kcdm^{-2} is observed at the south, which is about seven times greater than the minimum value. This observation proved that the sky luminance under a clear sky condition strongly depends on the solar position.

Partly cloudy sky

The partly cloudy sky can be regarded as the mean luminance distribution of all skies, excluding the clear and overcast skies [Narkamura et al., 1987]. Li and Lam [2001] reported that partly cloudy sky conditions prevailed in Hong Kong. An understanding of this prevailing sky condition, therefore, may form an important part of daylighting design processes. However, luminance distributions between the two extremes sky

conditions cannot be predicted accurately and easily. For instance, a small cloud in a clear sky condition could keep the sun obscured by slowly traversing across the sky while a small hole on the cloud in an overcast sky condition could remain open to the sun for a long time. Nevertheless, a partly cloudy sky was chosen from the database and the corresponding sky luminance distributions were presented in Figure 2.16. As expected, the pattern is quite complicated. The luminance ranges from 0.8kcdm^{-2} at the north to 14.1kcdm^{-2} at the south where is close to the solar position.

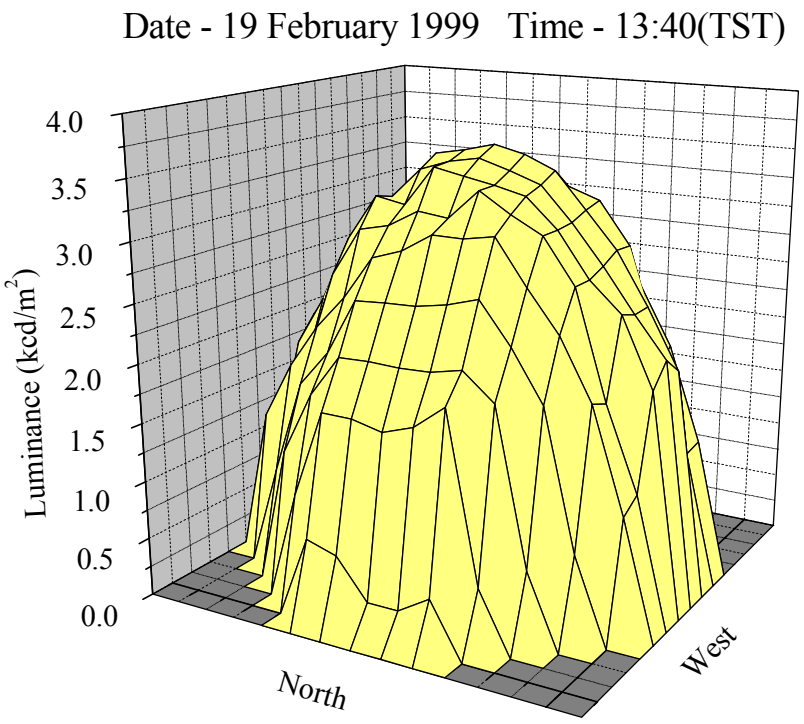

Figure 2.14 Measured sky luminance distribution for an overcast sky

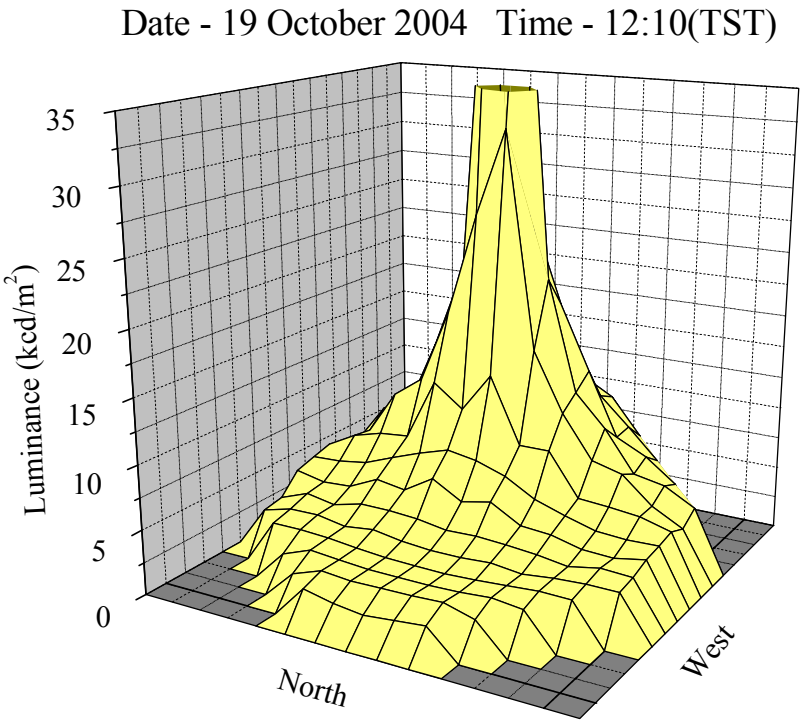

Figure 2.15 Measured sky luminance distribution for a clear sky

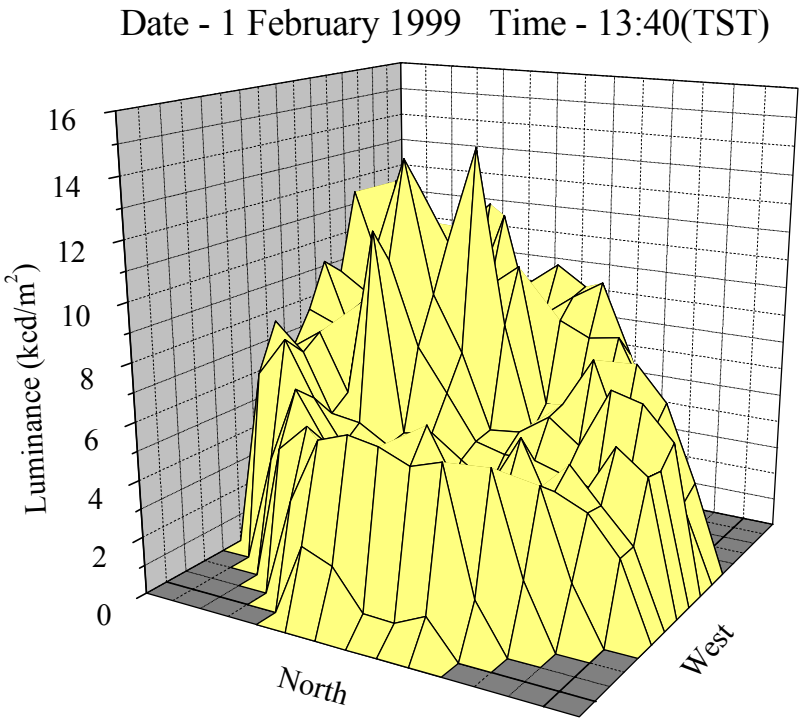

Figure 2.16 Measured sky luminance distribution for a partly cloudy sky

2.5 Summary

In this chapter, the daylight climate of Hong Kong has been studied by analyzing the quality controlled measurement data collected at City University of Hong Kong during 1999-2005. The peak horizontal G_v and D_v are normally occurred in summer and spring, respectively. Moreover, it has been observed that there is a strong relationship between the outdoor illuminance and solar position.

A brief introduction about the 3 general sky conditions has been mentioned. Their characteristics affect the selection of climatic parameters in the next chapter. Based on the 3 general sky conditions, chapter 3 will illustrate specifically the 15 standard skies adopted by CIE. To define the actual sky conditions in Hong Kong, the scanned sky luminance data will be used to determine the frequency distributions and the corresponding root mean square error (RMSE) of each standard sky type.

Chapter 3

CLIMATIC PARAMETERS AND CIE STANDARD SKY

An accurate estimation of daylight available outside the building is to acquire not just the total amount of external illuminance levels available but also the distributions of luminance across the sky vault [Li et al., 2006a]. The indoor daylight illuminance of a room is found to be related to the sky luminance distribution in the direction of view of the window [Rutten, 1994]. It is important to define sky luminance distribution under various conditions so that the daylight design and appraisal can be harmonized effectively [Darula et al., 2001].

CIE standard overcast sky

Moon and Spencer [1942] proposed a trigonometric function for calculating the luminance distribution of an overcast sky as shown in equation 3.1 and this model was adopted by the CIE as the standard overcast sky in 1955. The luminance distribution of an overcast sky is symmetrical about the zenith and changes with the elevation above the horizon. Most of the daylighting predictions in many parts of the world were

performed based upon the standard overcast sky.

$$\frac{L_\gamma}{L_Z} = \frac{1+2\sin\theta}{3} = \frac{1+2\cos Z}{3} \qquad (3.1)$$

where L_γ = sky luminance of a sky element (kcdm^{-2})

L_Z = sky luminance at the zenith (kcdm^{-2})

θ = altitude of a sky element (rad)

Z = zenith angle of a sky element (rad)

According to equation 3.1, it is interesting to note that the brightness of the zenith is three times brighter than the horizon. It has been pointed out in the literature that the CIE overcast sky performed the best among all worldwide models under a fully overcast sky and showed a good agreement with the sky luminance data measured in many parts of the world [Enarun and Littlefair, 1995].

CIE standard clear sky

Later, a clear sky luminance distribution function was developed by Kittler [1967] and was recommended as the CIE standard clear sky [CIE, 1973]. Similar with the standard overcast sky, the luminance distribution of the standard clear sky was expressed by the ratio of L_γ to L_Z as shown in equation 3.2. It should be noted that the luminance distributions are substantially affected by weather conditions and the solar positions [Littlerfair, 1994].

$$\frac{L_\gamma}{L_Z} = \frac{\left(1 - e^{-0.32/\sin\gamma}\right)\left(0.91 + 10e^{-3\chi} + 0.45\cos^2\chi\right)}{0.274\left(0.91 + 10e^{-3Z_s} + 0.45\cos^2 Z_s\right)} \quad (3.2)$$

where Z_s = zenith angle of the sun (rad)

χ = scattering angle, the shortest angular distance between the sun and a sky element (rad)

15 CIE standard skies

Nevertheless, the sky conditions between the overcast and clear sky cannot be determined satisfactorily by applying the two standard sky models. It is no longer satisfactory to determine the indoor daylight illuminance level with only two sky conditions [Kobav and Bizjak, 2003]. Therefore, a new set of standard skies should be developed to alleviate the problems. To standardize sky luminance distributions which occur in the nature, a total of three proposals were proffered and submitted to CIE for analysis.

Nakamura et al. [1985] introduced the luminance distributions of intermediate skies which represents all the sky conditions excluding the clear sky and overcast sky. The relative luminances of sky elements to the zenith luminance corresponding to α_s of every 10^o, from 10^o to 70^o were determined based on the long period measurement data.

By using statistical approach, Igawa et al. [1997] developed a set of twenty sky luminance distribution models based on the measured luminance data at Tokyo, which represents the relative patterns between the CIE standard clear and overcast sky conditions.

Kittler et al. [1997a] proposed a set of 15 sky types based on measured sky luminance distribution data in Berkeley (USA, 37.6^{o}N and 122.4^{o}W) during 1985 - 1986, Tokyo (Japan, 35.8^{o}N and 139.8^{o}E) and Sydney (Australia, 33.9^{o}S and 151.2^{o}E) in 1992. Scattering indicatrix and sky gradation function were combined to form a set of mathematical equations which can be regarded as the luminance distributions of the sky vault. The proposed set of sky types includes five overcast, five partly cloudy and five clear skies.

After voting in the CIE Technical Committee 3-15, the 15 sky types proposed by Kittler et al. [1997a] was recommended as the CIE standard skies [Darula et al., 2001]. The set of 15 standards skies adopted all existing CIE standards including CIE standard overcast and clear skies in a system covering the whole probable spectrum of skies in the world [Kittler et al., 1999]. This standard sky models are characterized by continuous mathematical expressions and changed smoothly in luminance from the horizon to the zenith and with the angular distance from the sun.

Tregenza [1999] analyzed these 15 standard skies. Using the recorded sky scans collected at four IDMP measuring stations, namely Singapore (1.5^{o}N and 104^{o}E), Fukuoka (Japan, 33.5^{o}N and 130.5^{o}E), Garston (UK, 51.7^{o}N and 0.4^{o}W) and Sheffield (UK, 53.5^{o}N and 1.5^{o}W), which represent tropical humid climate and temperate maritime climate. The author concluded that this standard set could provide a satisfactory overall framework for categorizing actual skies.

Apart from the above, measurements and studies of sky luminance distributions based on the 15 standard skies have been undertaken extensively in many parts of the world.

Li et al. [2003] studied the measurement data recorded in subtropical Hong Kong China (22.3^{o}N and 114.2^{o}E) against the 15 standard luminance distribution models. The authors found that the 15 standard skies were adequate to identify sky conditions in Hong Kong. Based on the three years hourly sky luminance records during 1999 to 2001, Li et al. [2003] concluded that a subset of five standard skies (i.e. sky types 1, 3, 6, 11 and 13) would be sufficient to describe Hong Kong sky conditions without any sufficient deterioration in the prediction accuracy. Later, Ng et al. [2007] studied further to understand the actual sky patterns of Hong Kong using 10-minute measurements during 2003 to 2005. The results indicated that a subset of 3 sky types (i.e. sky types 1, 8 and 13) is adequate to represent the sky conditions of subtropical Hong Kong.

Chirarattananon and Chaiwiwatworakul [2007] reported the results of characterization of the sky distributions under the standard sky models using the sky luminance data collected at topical north of Bangkok (Thailand, 14.08^{o}N, 100.62^{o}E) from 1999 to 2004. Shahriar and Mohit [2006] examined whether the proposed standard skies are applicable to Subang (Malaysia, 3.7^{o}N and 101.33^{o}E) and concluded that the standard skies are good representations of the sky conditions of Subang.

The actual sky condition is a critical factor for daylighting design. Once the skies have been interpreted, the basic outdoor illuminance at various inclined surfaces can be obtained for subsequent investigations. Nevertheless, sky luminance distributions particularly for the whole sky are scarcely available or exist for very limited periods. For places where such measured data are not obtainable, sky patterns should be interpreted using other sky indicators. Sky classification using common climatic indices is the first step to examine the standard skies in details.

The objectives of this chapter are i) to introduce the climatic parameters that will be used in this study. The frequency of occurrences of such sky indicators contributing the 3 typical sky conditions are established and presented. Their strengths and limitations for sky classification are discussed, and ii) to establish the occurrences of standard skies through comparison between the measured and calculated illuminance values for all 15 standard skies. Furthermore, the prevailing standard skies of Hong Kong are discussed and the 10-minute of sky luminance data recorded in the daytime of the period of 1999-2005 are used in the present analysis.

3.1 Climatic Variables under Typical Skies

Generally, sky conditions can be split into three typical categories. To identify the various sky conditions, different researchers used different climatic indices [Littlefair, 1994]. Many climatic parameters can be used for characterization of sky conditions. Three criteria should be considered to judge the merits and demerits of the parameters for sky interpretation i) accuracy – to minimize errors, ii) simplicity – such that the computation time can be kept to a minimum and manual checks are not prohibitively complex, and iii) parameters availability – so that sufficient data can be used. Instead of using solar radiation parameters that have been examined by various researchers, five common and synoptic meteorological variables regarding outdoor illuminance are introduced.

3.1.1 Solar altitude (α_s)

Sky luminance distributions for most sky types depend on the α_s in particular under clear skies. Under heavy overcast skies (sky standards 1, 3 and 5), the distributions of the sky luminance are symmetrical about the zenith and to change with the sky elevation above the horizon. It means that the sky luminance distributions for sky standards 1, 3 and 5 are independent of sky azimuth and solar location. For other sky standards, peak sky luminance would appear near to the solar position, decreasing with the distance from the sun. Solar altitude alone may not be a good factor to single out standard skies. However, a number of climatic parameters do vary with α_s and such features are useful for sky classification.

Solar altitude is the angular distance (between 0^o and 90^o) measured above the horizon in a vertical plan through the sun. It is the complement of solar zenith angle (Z_s) and has great effects for the solar energy reaching the earth surface. The higher the α_s the shorter distance that the solar radiation passing through the atmosphere and thus more solar energy arriving at the earth surface. α_s can either be measured or determined using the following equation [Kittler et al., 1992].

$$\sin \alpha_s = \sin\delta \sin L + \cos\delta \cos H \cos L \qquad (3.3)$$

where δ = solar declination angel (rad)

L = latitude of the measurement location (rad)

H = hour angle (rad)

3.1.2 Ratio of zenith luminance to horizontal diffuse illuminance (L_z/D_v)

The L_z/D_v ratio is one of the useful parameters to identify the sky conditions although they are far less commonly available and usually exist for very limited period of time [Karayel et al., 1984]. The integration of the luminance of each sky patch over the whole sky vault gives the horizontal illuminance. Since zenith is in a dominant position in the sky with respect to (i.e. normal to) the horizontal surface, D_v is substantially influenced by L_z. Therefore, a high probability of agreement (i.e. the zenith luminance is the major component contributing to D_v) does exist between these two climatic parameters.

According to the CIE sky standard, however, the L_z/D_v theoretical curves for the 15 standard skies are not parallel but they intersect with each other at the α_s of 35° or higher [Kittler et al., 1998]. Ambiguous results may be caused when using L_z/D_v ratio to identify the sky for place where high α_s dominates such as Hong Kong [Bartzokas et al., 2003; Markou et al., 2005]. The L_z/D_v ratio against α_s for the 15 different standard skies is shown in Figure 3.1.

The L_z/D_v ratios under the 3 typical sky conditions are computed and displayed in Figure 3.2. It can be seen that for overcast skies, the distribution is quite symmetrical with a peak L_z/D_v ratio at 0.39. Most of the readings appear between 0.24 and 0.54. As for partly cloudy days, the distribution is slightly shifted to left with a maximum L_z/D_v ratio occurring at 0.24. It spreads more widely than that in overcast skies, ranging from 0.09 to 1. Under clear sky conditions, a marked peak of 7% at L_z/D_v ratio of 0.15 is

observed. The distribution tends to cluster at L_z/D_v ratio between 0.11 and 0.29. All the features and the characteristics shown are in good agreement with the findings reported by Kittler et al. [1997a].

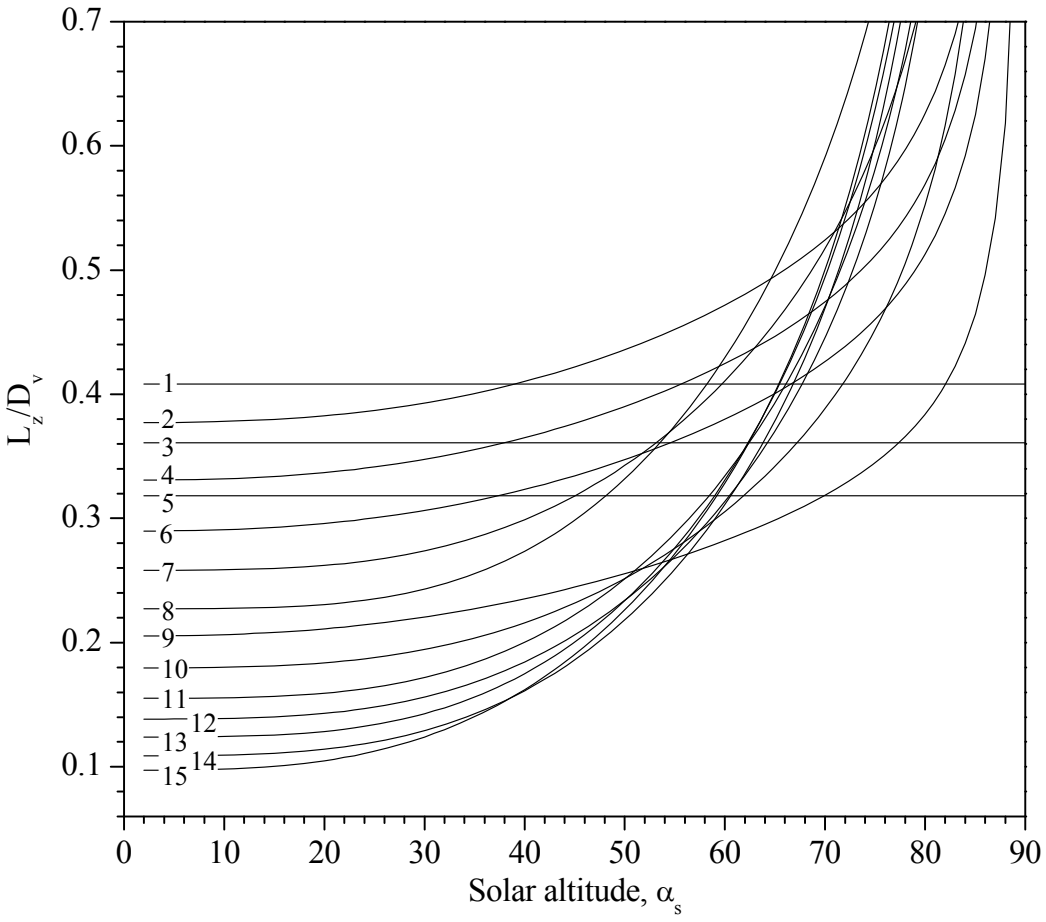

Figure 3.1 L_z/D_v plotted as a function of α_s for the 15 standard skies

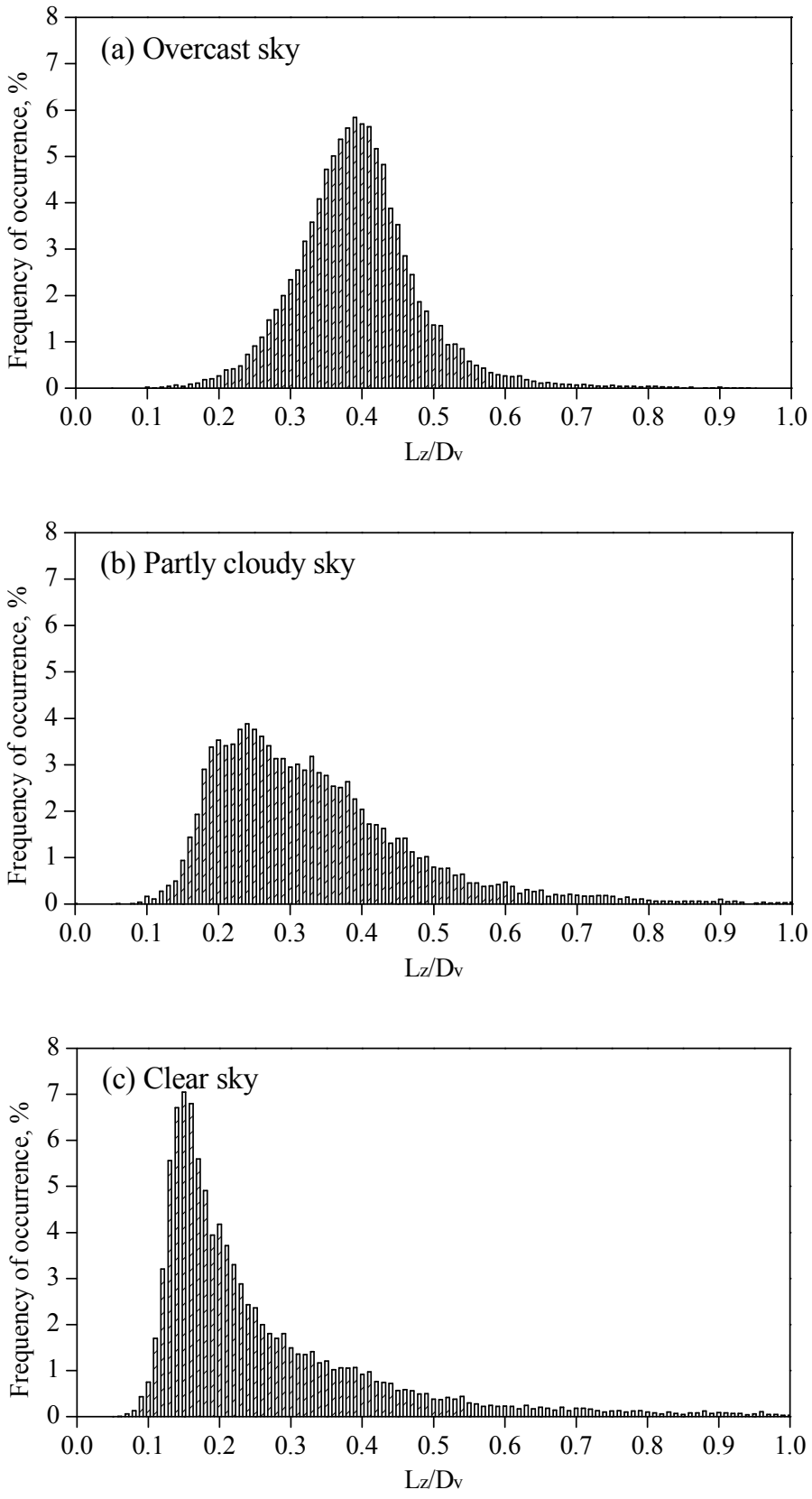

Figure 3.2 Frequency of occurrence of L_z/D_v for the 3 typical skies

3.1.3 Ratio of horizontal global illuminance to extraterrestrial illuminance (G_v/E_v)

The criteria for determining sky clearness using G_v usually take the form of G_v/E_v. At ground level, extraterrestrial illuminance (E_v) is reduced and split due to air attenuation, turbidity and clouds to form G_v. It should be noted that E_v on a horizontal surface can be determined as follows:

$$E_v = rE_{SC}\left(\cos L \cos\delta \cos H + \sin L \sin\delta\right) \tag{3.4}$$

where E_{SC} = solar constant (133.8klux)

r = eccentricity correction factor (dimensionless)

= $1.000110 + 0.034221 \cos\Gamma + 0.001280 \sin\Gamma + 0.000719 \cos 2\Gamma + 0.000077 \sin 2\Gamma$

Γ = day angle (dimensionless)

When the atmosphere is clear, a small fraction of outdoor illuminance is scattered, resulting in a predominant direct component with a large amount of G_v/E_v ratio. A large amount of outdoor illuminance is scattered for an overcast day, on the contrary, indicating a high portion of diffuse component with a low G_v/E_v ratio.

This ratio is a widely used index since it relies only on one measured parameter, global daylight illuminance which is recorded by many meteorological stations and the measurement is quite straightforward. Once G_v is obtained, the corresponding horizontal extraterrestrial illuminance can be determined by equation 3.4. However, G_v/E_v ratio is

neither zero nor unity. There is no clear-cut G_v/E_v ratio to indicate various sky conditions. Different values have been adopted by various researchers to show different sky conditions [Li and Lam, 2001; Li et al., 2004b].

Figure 3.3 presents the frequency distribution of G_v/E_v ratio at 0.01 intervals under the three typical sky conditions. It can be seen that G_v/E_v data tend to cluster more in the low range for overcast days and high range under clear skies. As for the partly cloudy skies, G_v/E_v data spreads widely on both low and high values. Peak values can be found at G_v/E_v ratios of 0.15, 0.39 and 0.68 for overcast, partly cloudy and clear skies respectively.

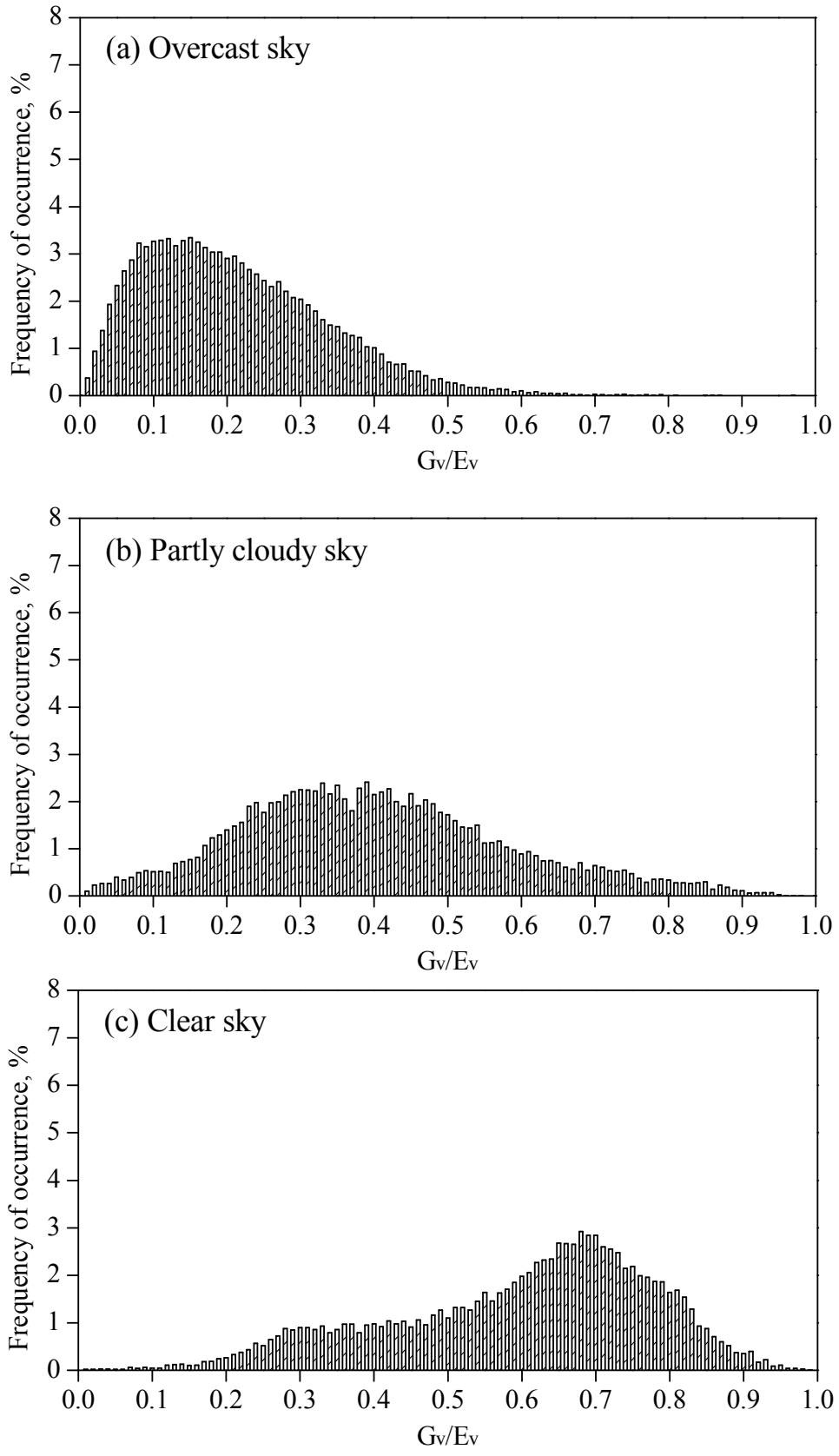

Figure 3.3 Frequency of occurrence of G_v/E_v for the 3 typical skies

3.1.4 Ratio of horizontal diffuse illuminance to extraterrestrial illuminance (D_v/E_v)

Similar with G_v/E_v ratio, D_v/E_v ratio can be used as a climatic index to interpret the sky conditions, particularly during overcast and cloudy situations with absence of the sun [Darula and Kittler, 2005]. D_v/E_v ratio depends on the diffuse illuminance which is measured using a shadow-ring to eliminate direct sunlight. Large D_v/E_v ratios often occur in partly cloudy sky. For low D_v/E_v ratios, they may indicate two extreme sky conditions (i.e. overcast skies and clear skies). Darula and Kittler [2004a; 2004b; 2005] analyzed the frequency of occurrence of D_v/E_v ratio using five years database recorded at Bratislava IDMP station during 1994 to 1998. A set of typical D_v/E_v ratio are suggested for different standard sky types as shown in Table 3.1.

The frequency of occurrence for D_v/E_v ratio under the 3 typical sky types was determined and is shown in Figure 3.4. It can be seen that the distribution is skewed to the left with a peak value of just over 4.3% at $D_v/E_v = 0.11$ and falls down to less than 0.1% at $D_v/E_v = 0.42$. The D_v/E_v ratios of 0.5 or more are rare to none. Under a partly cloudy sky, the increases from 0.1% at $D_v/E_v = 0.01$ to the peak 5.4% at $D_v/E_v = 0.29$ and drops rapidly to 0.2% at $D_v/E_v = 0.42$. As the sky becomes clear, the distribution is quite symmetrical at $D_v/E_v = 0.24$ with 6.8%, ranging from $D_v/E_v = 0.09$ to $D_v/E_v = 0.41$.

Table 3.1 Recommended D_v/E_v ratio for the 15 standard skies

No	Gradation group	Indicatrix group	Type of sky	Typical D_v/E_v
1	I	1	CIE Standard Overcast Sky, Steep luminance gradation towards zenith, azimuthal uniformity	0.10
2	I	2	Overcast, with steep luminance gradation and slight brightening towards the sun	0.10
3	II	1	Overcast, moderately graded with azimuthal uniformity	0.15
4	II	2	Overcast, moderately graded and slight brightening towards the sun	0.20
5	III	1	Sky of uniform luminance	0.22
6	III	2	Partly cloudy sky, no gradation towards zenith, slight brightening towards the sun	0.30
7	III	3	Partly cloudy sky, no gradation towards zenith, brighter circumsolar region	0.35
8	III	4	Partly cloudy sky, no gradation towards zenith, distinct solar corona	0.40
9	IV	2	Partly cloudy, with the obscured sun	0.35
10	IV	3	Partly cloudy, with brighter circumsolar region	0.30
11	IV	4	White – blue sky with distinct solar corona	*0.26 #0.30
12	V	4	CIE Standard Clear Sky, low luminance turbidity	*0.25 #0.30
13	V	5	CIE Standard Clear Sky, polluted atmosphere	*0.26 #0.30
14	VI	5	Cloudless turbid sky with broad solar corona	*0.28 #0.30
15	VI	6	White – blue turbid sky with broad solar corona	*0.28 #0.30

Note: * sunny situations, # situations with sun shaded

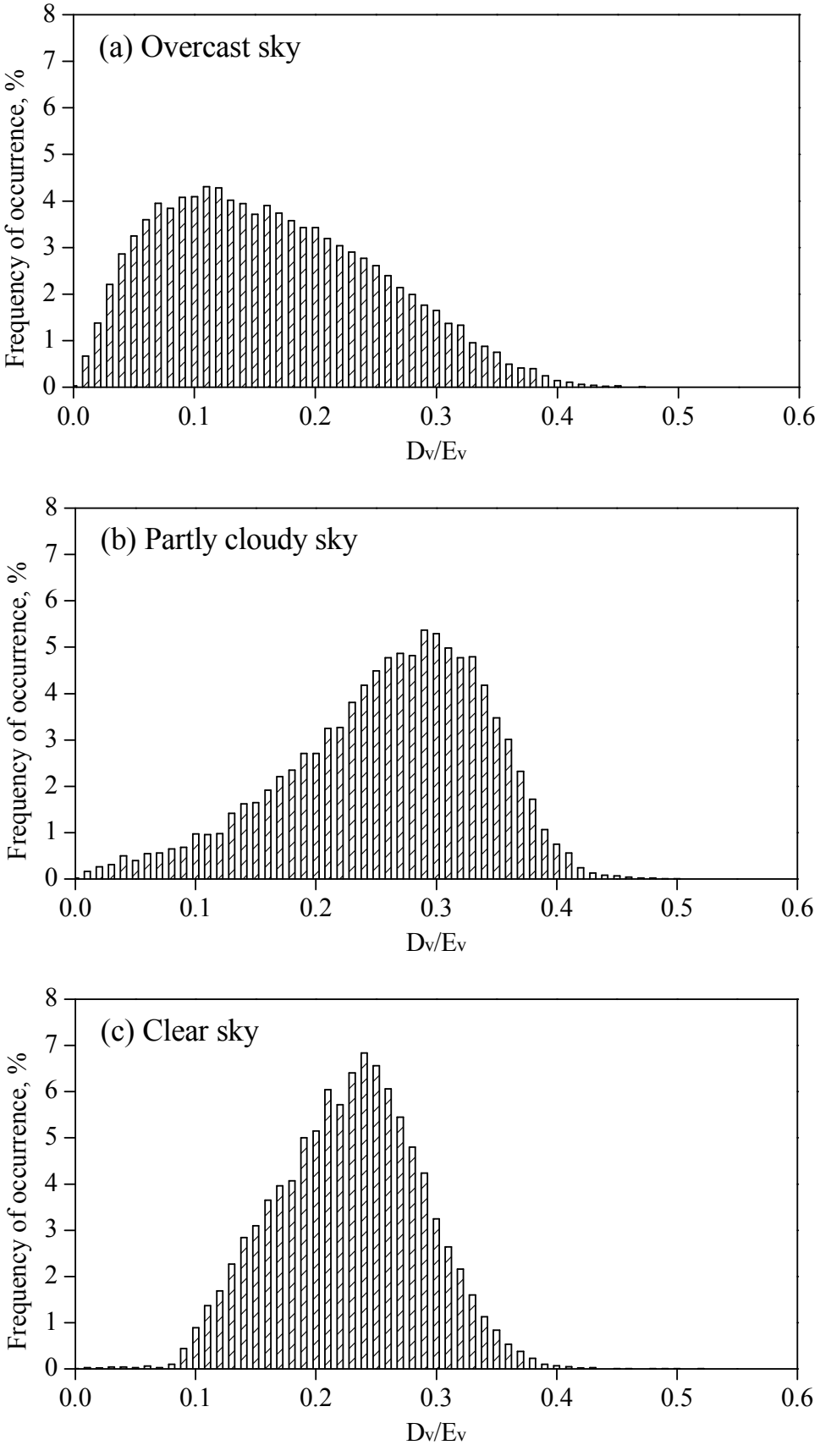

Figure 3.4 Frequency of occurrence of D_v/E_v for the 3 typical skies

3.1.5 Luminous turbidity (T_v)

The atmosphere consists of different substances such as water vapour, dust and aerosols. Due to the effect of absorption and scattering of light, the amount of daylight illuminance to the ground through the atmosphere is reduced. The actual value of T_v represents the number of absolutely clean atmospheres substituting a real case by having the same $(G_v\text{-}D_v)/E_v$ value under the actual α_s and directional air mass [Kittler and Darula,1998; Soler and Gopinathan, 2000]. The lowest T_v is unity while the highest is infinity when $(G_v\text{-}D_v)/E_v = 0$ in the direction of sun beams.

Darula and Kittler [2005] reported that luminous turbidity is the most appropriate descriptor to roughly classify all clear sky types since T_v value is independent of the α_s changes, its daily course is practically stable except for changes caused by higher water evaporation or cloudiness rise. The usual T_v values are between 10 and 20 under partly cloudy sky types and below 10 for clear sky types. Various T_v criteria can be applied to classify individual partly cloudy and clear skies [Li and Lau, 2007; Li and Tang, 2008]. The general formula for calculating T_v can be expressed as follows [Kittler et al., 1998]:

$$T_v = -\frac{\ln(G_v - D_v / E_v)}{a_v m_i} \tag{3.5}$$

The luminous extinction coefficient, a_v proposed by Navvab et al. [1984] can be calculated as:

$$a_v = \frac{1}{9.9 + 0.043 m_i} \tag{3.6}$$

The optical air mass, m_i from Kasten and Young [1989] as given by:

$$m_i = \frac{1}{\sin\alpha_s + 0.50572(\alpha_s + 6.07995)^{-1.6364}} \tag{3.7}$$

The frequency distribution of T_v at unity interval under the three general sky conditions is presented in Figure 3.5. It can be seen that T_v spreads widely on both low and high values under overcast days, ranging from 0.1% to 3.1%. As for the partly cloudy skies, T_v data tend to cluster more in the low range. Under clear sky conditions, a marked peak of 16% at $T_v = 6$ is observed. The distribution tends to cluster at T_v between 3 and 12.

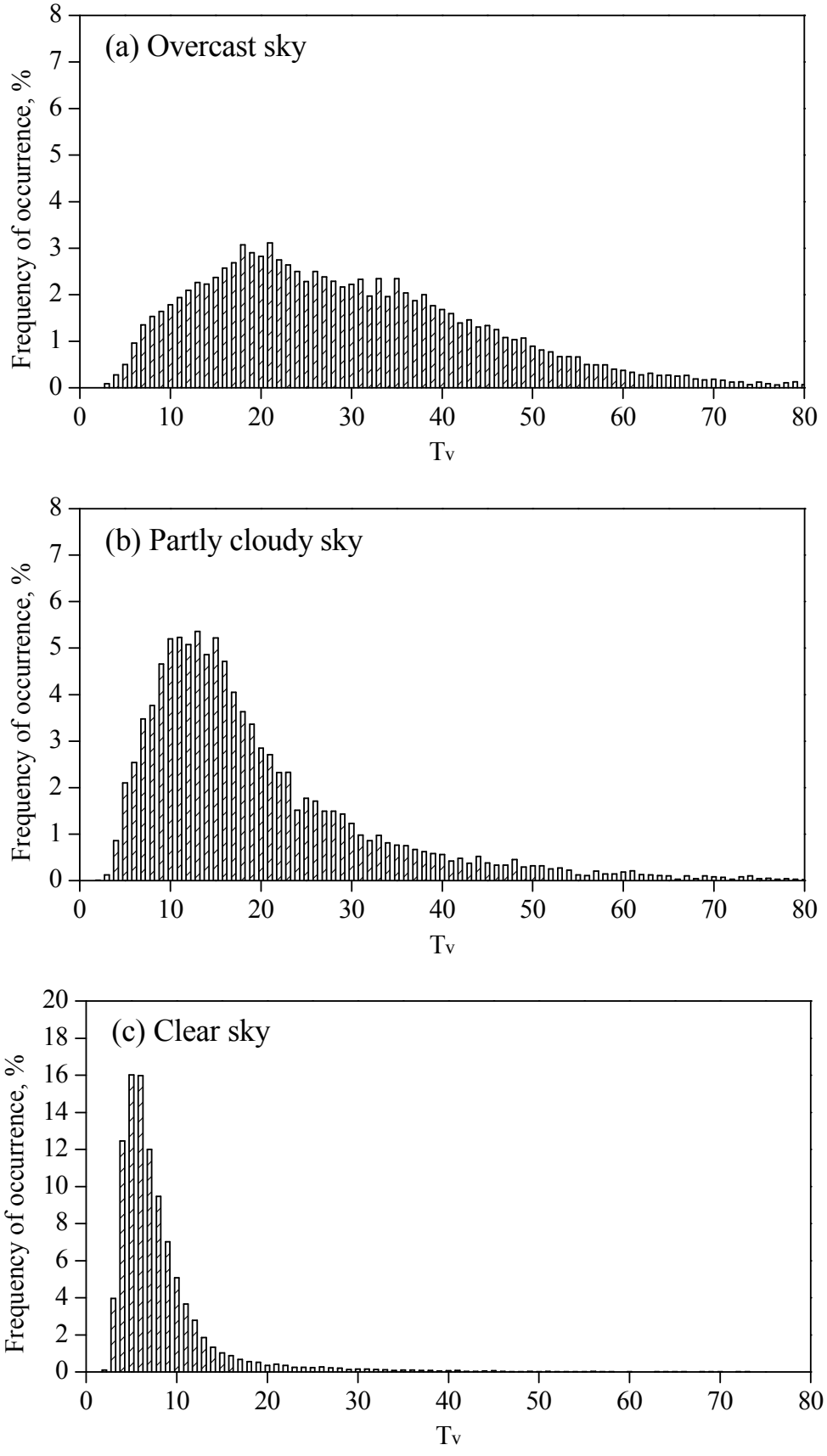

Figure 3.5 Frequency of occurrence of T_v for the 3 typical skies

3.2 CIE Standard Skies

The set of the 15 standard skies includes the CIE clear sky distribution, a uniform luminance distribution, and a close approximation to the CIE overcast sky. In general, the standard skies contain 5 overcast, 5 partly cloudy and 5 clear sky types covering the whole probable spectrum of skies in the world [Kittler et al., 1999]. The distributions are described by continuous mathematical expressions that change smoothly in luminance from the horizon to zenith and with the angular distance from the sun. The standard formula defining the relative sky luminance, l_v, on any standard sky can be considered as a combination of the gradation $\varphi(Z)$ and the indicatrix function $f(\chi)$. The 15 standard relative luminance distributions which are based on 6 groups of a and b values for the gradation function and 6 groups of c, d and e values for the indicatrix function.

$$l_v = \frac{L_\gamma}{L_Z} = \frac{f(\chi)\varphi(Z)}{f(Z_S)\varphi(0°)} \tag{3.8}$$

The standardized gradation is defined by appropriate a and b variables as:

$$\frac{\varphi(Z)}{\varphi(0°)} = \frac{1 + a\,\exp(b/\cos Z)}{1 + a\,\exp b} \tag{3.9}$$

The relative scattering indicatrix function can be modeled by an exponential function with adjustable coefficients c, d and e as:

$$\frac{f(\chi)}{f(Z_S)} = \frac{1 + c\left[\exp(d\chi) - \exp(d\pi/2)\right] + e\cos^2\chi}{1 + c\left[\exp(dZ_S) - \exp(d\pi/2)\right] + e\cos^2 Z_S} \tag{3.10}$$

The exponential term exp($d\chi$) represents the effect of Mie scattering, which decreases rapidly with distance from the sun. The $\cos^2\chi$ term is due to the Rayleigh scattering and is zero at 90° to the direction of the sun [Littlefair, 1994]. Both gradation and indicatrix functions are of 6 types covering the usual range of homogeneous cases from heavy overcast to cloudless skies. The combinations can form a large number of skies but only 15 relevant types were selected to be the standard set. Table 3.2 shows the details of the 15 standard skies [Kittler et al., 1997a].

Table 3.2 A set of 15 sky standards and their parameterization

No (Code)	Type of sky	for gradation		for indicatrix		
		a	b	c	d	e
1 (I1)	CIE Standard Overcast Sky, Steep luminance gradation towards zenith, azimuthal uniformity	4	-0.7	0	-1	0
2 (I2)	Overcast, with steep luminance gradation and slight brightening towards the sun	4	-0.7	2	-1.5	0.15
3 (II1)	Overcast, moderately graded with azimuthal uniformity	1.1	-0.8	0	-1	0
4 (II2)	Overcast, moderately graded and slight brightening towards the sun	1.1	-0.8	2	-1.5	0.15
5 (III1)	Sky of uniform luminance	0	-1	0	-1	0
6 (III2)	Partly cloudy sky, no gradation towards zenith, slight brightening towards the sun	0	-1	2	-1.5	0.15
7 (III3)	Partly cloudy sky, no gradation towards zenith, brighter circumsolar region	0	-1	5	-2.5	0.3
8 (III4)	Partly cloudy sky, no gradation towards zenith, distinct solar corona	0	-1	10	-3	0.45
9 (IV2)	Partly cloudy, with the obscured sun	-1	-0.55	2	-1.5	0.15
10 (IV3)	Partly cloudy, with brighter circumsolar region	-1	-0.55	5	-2.5	0.3
11 (IV4)	White – blue sky with distinct solar corona	-1	-0.55	10	-3	0.45
12 (V4)	CIE Standard Clear Sky, low luminance turbidity	-1	-0.32	10	-3	0.45
13 (V5)	CIE Standard Clear Sky, polluted atmosphere	-1	-0.32	16	-3	0.3
14 (VI5)	Cloudless turbid sky with broad solar corona	-1	-0.15	16	-3	0.3
15 (VI6)	White – blue turbid sky with broad solar corona	-1	-0.15	24	-2.8	0.15

3.3 Standard Sky Selection

Tregenza [1999] developed a statistical approach to determine the standard distribution that is best-fit to each measured sky distribution. The author proposed a representative room of a building and compared the room illuminances calculated using the best-fit standard with those calculated from the actual sky. The best-fit standard sky selected is the one with the lowest RMSE. Later, Tregenza [2004] modified his approach which does not depend on the selection of the representative room. The calculations were simplified and became more general. In the present study, this best-fitting approach proposed by Tregenza [2004] was used to define the sky conditions in Hong Kong.

In general, a subset of the CIE standard skies including overcast, partly cloudy and clear sky conditions are sufficient to describe the daylight climates at any particular site since some standard skies are rarely applicable. [Li et al., 2003; Ng et al., 2007; Tregenza, 2004]. The selection of the sky luminance models representing overcast, partly cloudy and clear sky was examined.

To identify the set of standard skies in details, the luminance distributions of individual standard skies were modeled and compared with the scanned sky luminance readings. The modeled sky luminance was normalized to the horizontal diffuse illuminance by multiplying all the luminance values with the normalization ratio (NR) as:

$$NR = \frac{\Sigma L_{mea} \cos\theta \sin\theta \, d\theta \, d\beta}{\Sigma l_{pred} \cos\theta \sin\theta \, d\theta \, d\beta} \tag{3.11}$$

where L_{mea} = measured sky point luminance (kcdm^{-2})

l_{pred} = predicted sky point luminance in relative form (dimensionless)

β = azimuth of a sky patch (rad)

The formula (i.e. Equation 3.8) to determine the sky luminance is in relative form and therefore, l_{pred} should be used in the denominator of Equation 3.11. The NR is not the same as that in Littlefair's paper [1994] which the models for analysis were mainly based on absolute sky luminance. The relative (not absolute value) sky luminance for each sky patch was determined using Equation 3.8. Once the integrated diffuse illuminance and l_{pred} at each sky point has been obtained, the predicted sky luminance (L_{pred}) which is the product of NR and l_{pred} can be easily computed. An alternative would be to divide all sky luminance readings by the zenith luminance, but this can cause huge measuring error when the sun is near to the zenith [Tregenza, 2004]. For low-latitude climates (e.g. Hong Kong) when the sun is frequently within a small angular distance from zenith, a normalization with respect to the diffuse horizontal illuminance would, therefore, be more appropriate. Once normalized, the performance of each standard sky luminance model was assessed by the RMSE. The standard sky selected is the one with the lowest RMSE.

$$RMSE = \sqrt{\frac{1}{N}\sum\left(\frac{L_{pred} - L_{mea}}{L_{mea}}\right)^2} \tag{3.12}$$

where N = number of readings (dimensionless)

3.3.1 Results of complete set of standard skies

The frequency of occurrence of the 15 standard skies using the best-fitting approach (a), and the corresponding RMSE based on the complete standard set (b) are presented in Figure 3.6. Large variations can be observed for individual sky types and a detailed analysis of Figure 3.6 revealed that the overcast and clear skies (i.e. overcast – sky numbers 1 to 5 and clear – sky numbers 11 to 15) represent about 78% of the Hong Kong sky conditions. The intermediate skies (i.e. sky numbers 6-10) account for the remaining 22%. Sky numbers 1 and 3 dominate in the overcast sky whereas sky numbers 6, 7 and 8 appear more frequently when compare with the other partly cloudy skies. For clear sky condition, sky number 13 is the main sky type in Hong Kong. The RMSE for all the 15 standard skies is 23.8% using best-fitting approach.

Figure 3.6 presents the best-fit standard skies with high frequency of occurrence values give in low RMSE results and vice-versa. For instance, sky numbers 2, 5 and 9 have frequency of occurrence of less than 5% with the corresponding RMSEs of over 32%. With the frequency of occurrence over 12.7%, the RMSEs for sky numbers 1, 3 and 13 are less than 24%. This indicates that some standard skies appear infrequently and their exclusion can simplify the analysis without significantly lowering the overall accuracy.

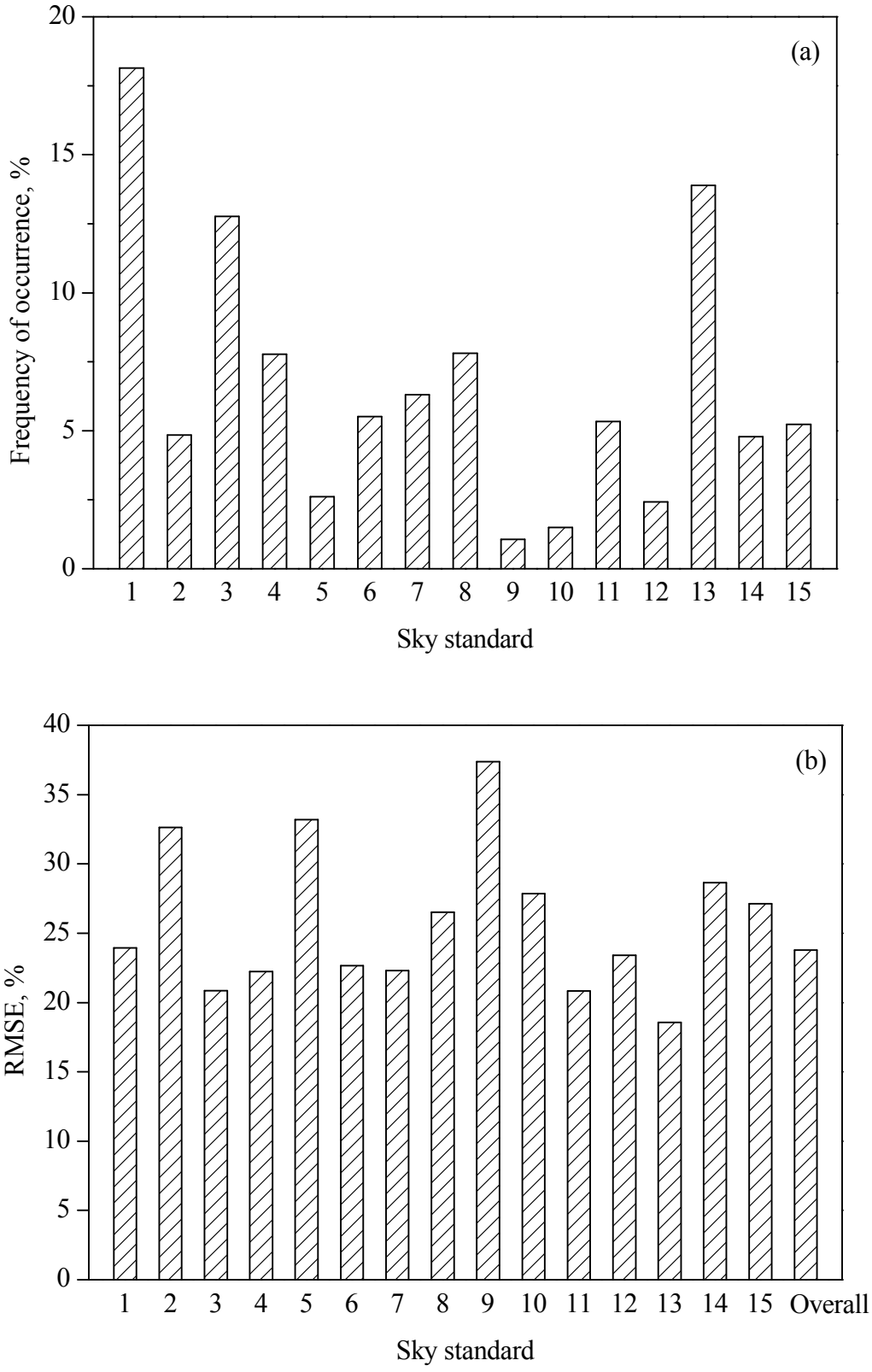

Figure 3.6 (a) Frequency of occurrence for the best-fit 15 standard skies, (b) RMSE of the luminance distribution of actual skies to the best-fit 15 standard distributions

3.3.2 Results of prevailing set of standard skies

Individual standard skies with high frequency of occurrence and low RMSE are appropriate to represent the prevailing sky conditions for a given location. In this regard, the ratio of frequency of occurrence to RMSE for the 15 best-fit standard skies was considered and Figure 3.7 presents the results. It can be seen that sky numbers 1 and 3 of overcast conditions have ratios over 0.6, which is far more than the other 3 overcast sky types (i.e. sky numbers 2, 4 and 5). For partly cloudy conditions, the ratio for sky numbers 6, 7 and 8 are between 0.24 and 0.3 but those for sky numbers 9 and 10 is close to zero. Referring to clear sky type, sky number 13 has the highest ratio of over 0.75. Based on these findings, sky numbers 1, 3, 6, 7, 8 and 13 were selected to represent the prevailing sky standards in Hong Kong.

Similar with section 3.3.1, the same database and approach were applied to determine the standard sky distributions. However, the outcomes of the calculation are 6 prevailing skies rather than the 15 standard skies. The remaining standard skies were discarded systematically by comparing the RMSE of the best-fit standard sky. When an excluded sky type is the best-fit to a particular scan, the sky type giving the next best-fit to that scan was substituted. The steps were repeated until the 6 prevailing skies were obtained. The frequency of occurrence and the RMSEs for the 6 prevailing skies were assessed and graphical presentations were used as shown in Figure 3.8. The largest frequency of occurrence of 29.3% was found in sky number 13 which has been selected to be the most representative clear skies from the CIE standard skies in Hong Kong. This is not surprising, given that Hong Kong is currently infamous for its poor air quality and polluted conditions. Comparing with the complete set of 15 standard skies, the frequency of occurrence for all partly cloudy skies are increased (sky number 6 – 12.5%,

sky number 7 – 9.5% and sky number 8 – 10.8%). For sky standards 1 and 3 the frequency of occurrence varies slightly. The RMSEs range from 23.6% for sky standard 13 to 27.9% for sky standard 8. The overall RMSE for the reduced set of the 6 prevailing sky standards in Hong Kong is 25% which is 1.2% greater than the complete set of 15 standard skies.

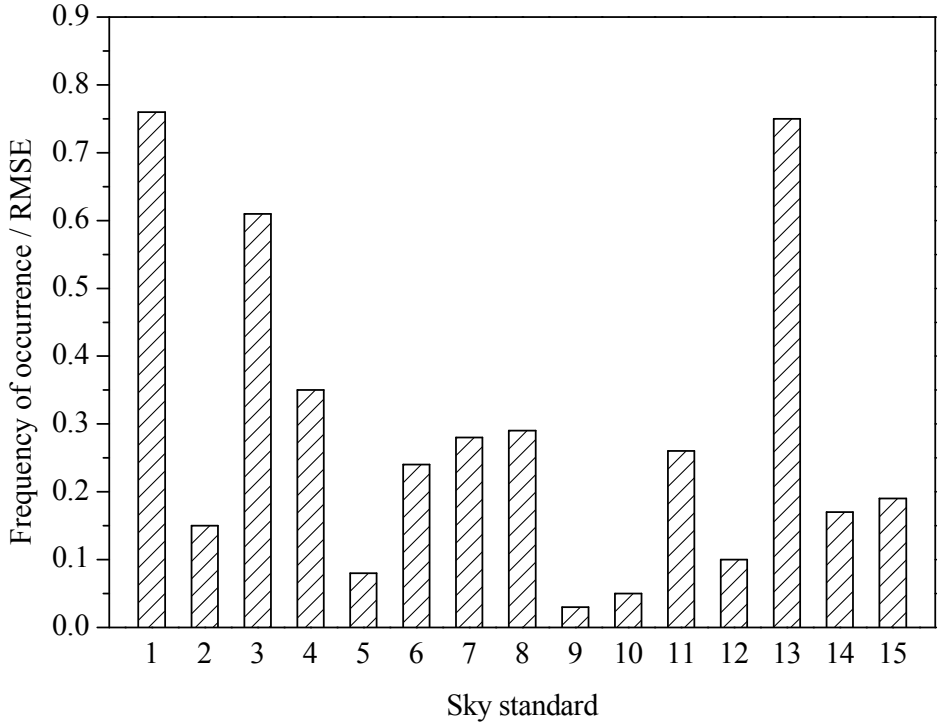

Figure 3.7 The ratios of frequency of occurrence to RMSE for the 15 standard skies

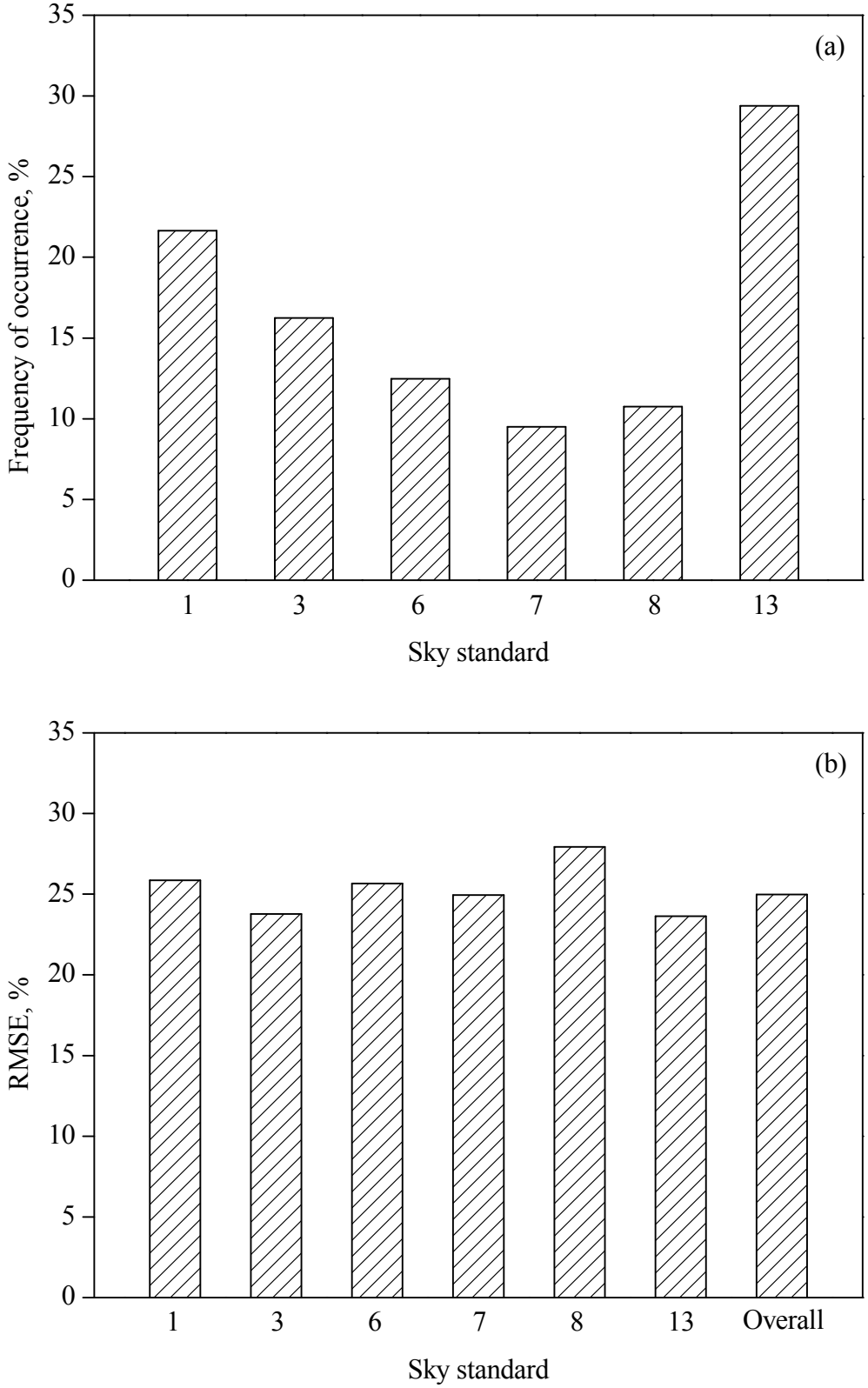

Figure 3.8 (a) Frequency of occurrence for the best-fit 6 prevailing standard skies, (b) RMSE of the luminance distribution of actual skies to the best-fit 6 prevailing standard skies

3.3.3 Results of subset of 3 standard skies

To further simplify the sky models, the overall RMSE against the number of standard skies available was presented in Figure 3.9. Similar with the results reported by Tregenza [2004], Li et al. [2003] and Ng et al. [2007], the overall RMSE does not change sharply as the number of sky type available is reduced to three. If 13 out of 15 of the sky distributions are eliminated, however, the overall RMSE increases significantly to 29.2%. The results suggested that a subset of three standard sky types is adequate to represent the sky conditions of Hong Kong. To simplify the analysis, three standard skies with the highest value of the ratio of frequency of occurrence to RMSE as shown in Figure 3.7 representing overcast, partly cloudy and clear types are selected for the study. The representative sky types for the 3 general sky conditions are sky numbers 1, 8 and 13.

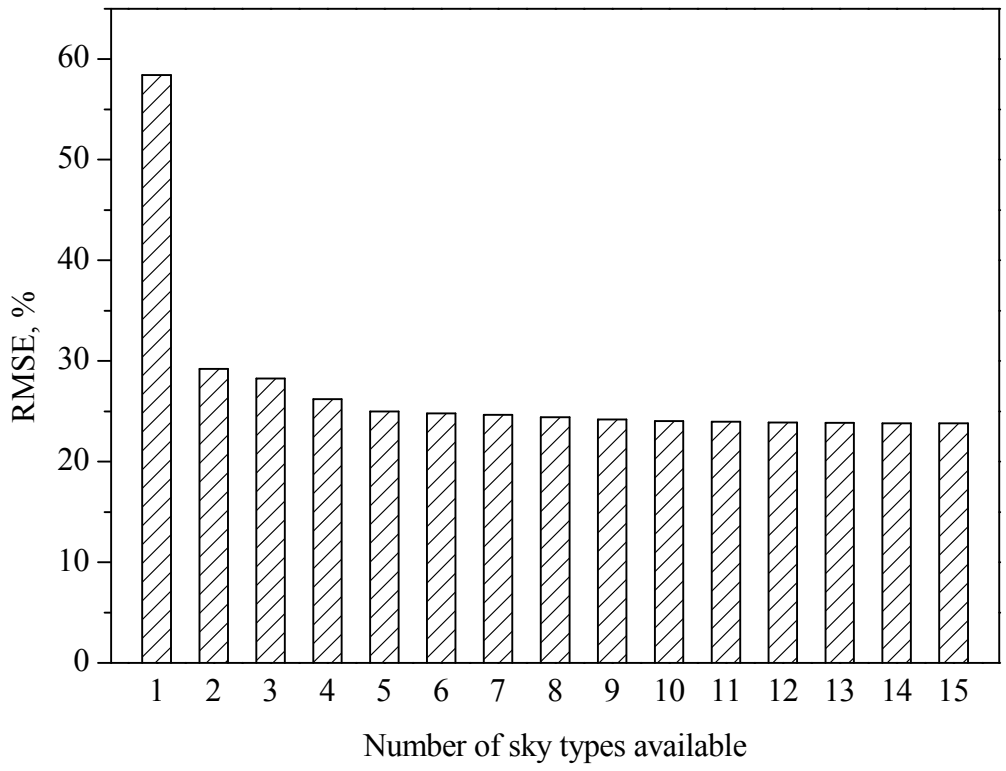

Figure 3.9 Overall RMSE against number of sky types available

Again, the frequency of occurrence and the RMSEs for these 3 skies were determined and the results are given in Figure 3.10. It can be seen that both the frequency of occurrence and RMSE values increase when all the sky conditions were interpreted by 3 standard sky distributions. Overcast sky (i.e. sky number 1) and clear sky (i.e. sky number 13) contribute about 80% of Hong Kong sky condition. The frequency distributions of the best-fit 3 standard skies are similar to the study reported by Ng et al. [2007] using sky luminance scans collected in Hong Kong. As expected, limiting the sky conditions to three skies could simplify the sky classification without a significant loss of overall accuracy. The overall RMSE for the subset of 3 standard skies is 27% which is 3.2 and 2% greater than the findings obtained from the complete set of 15 standard skies and the prevailing set of 6 standard skies, respectively.

3.4 Summary

This chapter was a first step to develop and evaluate the sky classification approaches proposed in the next two chapters. One of the aims of chapter 3 was to select the appropriate climatic parameters to interpret the various sky conditions. A total of five common climatic parameters (i.e. α_s, L_v/D_v, G_v/E_v, D_v/E_v and T_v) have been suggested and their frequency distributions were presented graphically under the 3 typical sky categories. Also, the luminance distributions of each sky standard were determined and compared with the scanned sky luminance readings to identify the sky types for subtropical Hong Kong. Once the sky types have been classified, the corresponding climatic parameters can be used to evaluate the sky classification approaches proposed in chapters 4 and 5.

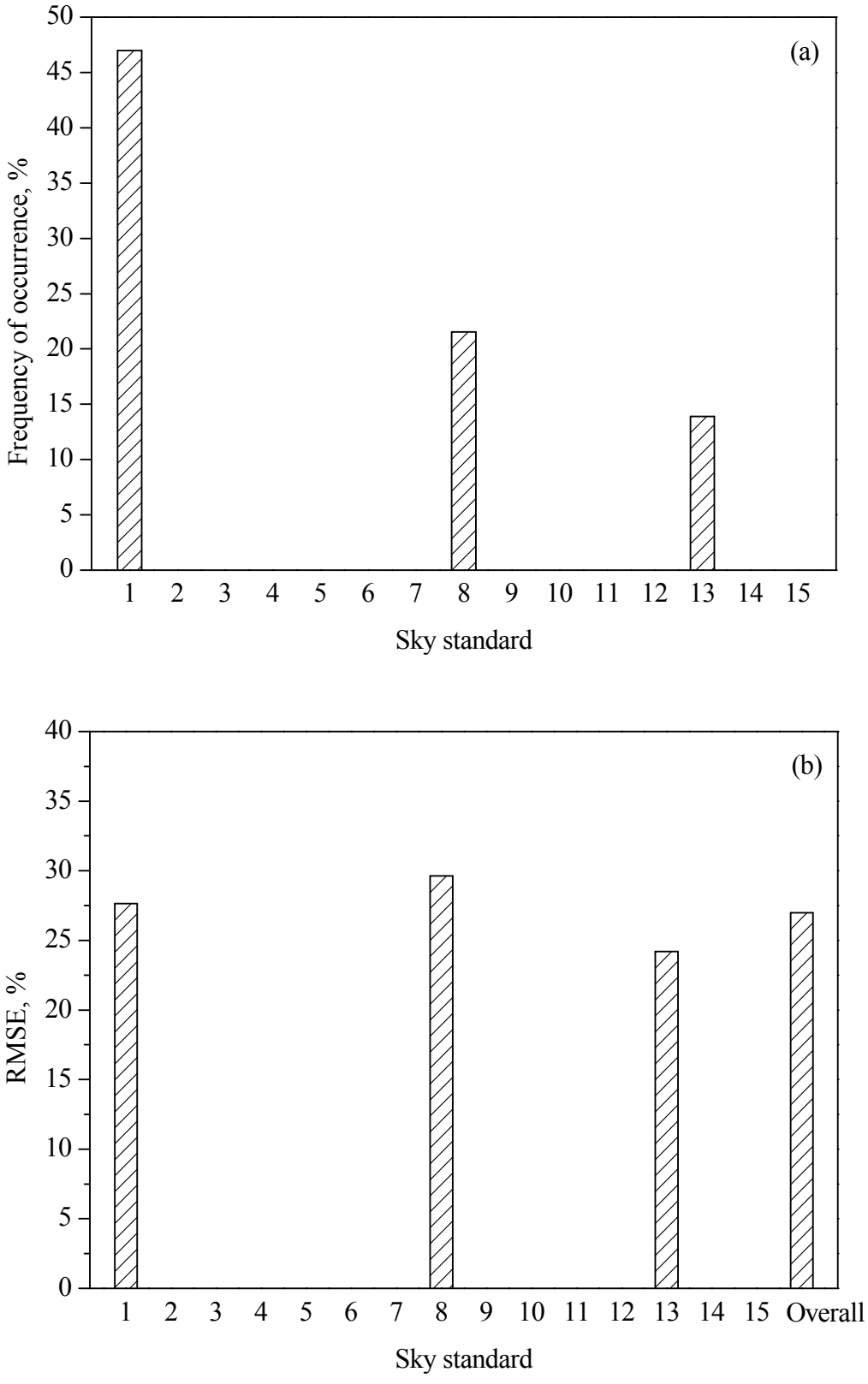

Figure 3.10 (a) Frequency of occurrence for the best-fit 3 standard skies, (b) RMSE of the luminance distribution of actual skies to the best-fit 3 standard skies

Chapter 4

SKY CLASSIFICATION I:

CLIMATIC PARAMETERS

The sky luminance distribution is crucial for designing an energy-efficient scheme for buildings [Wittkopf et al., 2006], determining the daylight illuminance level inside a room [Li, 2007] and evaluating the performance of daylighting software [Bellia et al., 2000]. In many part of the world, however, sky luminance measurements are still not available. In interpreting sky conditions, climatic parameters are often used as weighting factors to indicate the degree of sky clearness such that sky luminance distributions can then be classified. To identify the various sky conditions, different researchers used different climatic indices [Littlefair, 1994].

As mentioned, L_z/D_v ratio can characterize the momentary sky brightness and theoretically can classify the measurement into one of the 15 standard sky patterns. The change of L_z/D_v ratio only depends upon α_s excluding the overcast skies with unity indicatrix [Kittler et al., 1998; Kittler and Darula, 1998; Kittler and Darula, 2000]. Later, Kittler et al. [1999] reported that G_v or P_v together with simultaneous D_v is essential for

the sky classification under arbitrary cloudy sky conditions. Kittler and Darula [2002] suggested various parameters to interpret the sky conditions such as T_v and D_v/E_v ratio. There are a number of appropriate climatic variables to characterize the daylight climates. They concluded that G_v/E_v ratio, together with L_z/D_v ratio is sufficient to identify the typical skies.

Li and Lam [2001] identified the prevailing sky conditions in Hong Kong using K_t, K, cloud cover (CLD) and sunshine hour (SH). The frequency of occurrences and cumulative frequency distributions of these four parameters have been investigated and presented in graphical approaches. Three ranges of each parameter based on the frequency distributions of individual variables appearing in each typical sky condition are suggested to represent the 3 sky types. Later, Li et al. [2004b] suggested new criteria to improve the sky classification results. It has been pointed out that more accurate results can be obtained when applying two parameters (CLD and SH) simultaneously as the indicators to represent the 3 typical skies. Further studies about identification of the sky conditions particularly in overcast and non-overcast sky conditions can be found in the literature. [Li et al., 2004a; Li and Lau, 2007]

Younes and Muneer [2007] gathered the daylight measurements from six sites, namely Brachnell (UK, 51.26^{o}N and 0.45^{o}E), Gerona (Spain, 41.97^{o}N and 2.88^{o}N), Madrid (Spain, 40.45^{o}N and 3.73^{o}W), Chennai (India, 13.0^{o}N and 80.18^{o}E), Mumbai (19.12^{o}N and 72.85^{o}E) and Pune (India, 18.53^{o}N and 73.85^{o}E). Based on the different combinations of the meteorological indicators, totally nine clear sky tests have been proposed and evaluated independently by the six datasets. The results indicated that the clear sky test based on K_t, K, CLD and Linke turbidity (T_{LK}) is the most appropriate

way for classifying cloudless and low turbid skies.

Instead of using climatic parameters, Markou et al. [2007] classified the sky luminance distributions in Garston (UK, 51.71^{o}N and 0.38^{o}W) by multivariate statistical methods of factor analysis and cluster analysis. By using the outdoor illuminance and sky luminance measurements, the proposed method gave almost the same results at low α_s when comparing with the L_z/D_v sky classification technique. The difference of the results increases, however, as α_s increases.

This chapter presents the statistical approach on the classification of standard sky luminance distributions using various appropriate climatic parameters. A comparative assessment of the Bartzokas et al. [2003; 2005] and the proposed methods against scanned sky luminance data collected during 1999 and 2005 is reported. Characteristics of the findings are elaborated and discussed.

4.1 Bartzokas et al. Approach

Bartzokas et al. [2003; 2005] modified the L_z/D_v sky classification method suggested by Kittler et al. [1998] to study the daylight climates at two cities in Europe, namely Athens and Bratislava. The authors reported that the prevailing sky types for Bratislava are the overcast sky with steep luminance gradation and slight brightening towards the sun in winter and the white-blue sky with a distinct solar corona in summer. The most frequent sky types for Athens are the cloudless polluted sky in both winter and summer. Markou et al. [2005] further validated the methodology established by Bartzokas et al.

[2003; 2005] using half-hourly data recorded in Sheffield (UK, 53.4°N and 1.5°W) during winter of 1993 to 1994. Sky numbers 2, 6 and 13 are the most frequent sky types when examining the 3 general sky conditions separately, with a frequency of 19.5%, 7.0% and 7.6% respectively. The sky classification approach and the methodology proposed by Bartzokas et al. [2003; 2005] are briefly introduced in this section. Ten-minute climatic data collected in Hong Kong are used to evaluate this approach and the major findings are presented.

4.1.1 Methodology

This approach is based on the theoretical L_z/D_v curves of the 15 sky standards as shown in Figure 3.1. However, it can be seen that the theoretical L_z/D_v curves converge and they overlap at α_s of 35° or above. This could lead to ambiguous results in sky classification for high α_s. To overcome this difficulty, Bartzokas et al. established the following procedure. For each set of data, the L_z/D_v ratios at a particular α_s is initially computed based on measurements of the individual L_z and D_v parameters and compared with the L_z/D_v ratios of the 15 theoretical curves. Each observation is allocated to one of the 15 standard skies only when the measured L_z/D_v ratio lies within a zone of ±2.5% around the selected theoretical curve. Due to convergence and intersection of the theoretical curves, one observation may be interpreted in more than one standard sky. In such a case, G_v/E_v ratio for the specific observation is further compared to the average G_v/E_v ratios estimated from all observations at the same α_s for a particular standard sky. The observation is then classified in the standard sky with the nearest value.

Bartzokas et al. also defined two terms, namely ‘normal’ and ‘opposite’ cases. Of the 15

standard skies, the first six (sky numbers 1 to 6) represent overcast conditions associated with no direct sunlight and the direct-normal solar irradiance (I_{nB}) must be less than 120Wm^{-2}. To the contrary, for the remaining nine sky types (sky numbers 7 to 15) the condition $I_{nB} \geq 120 Wm^{-2}$ is valid. All these are called normal cases. When the observations are not categorized as normal cases, the classification is repeated under 'opposite' conditions (i.e. $I_{nB} \geq 120 Wm^{-2}$ for sky numbers 1 to 6 and $I_{nB} < 120 Wm^{-2}$ for sky numbers 7 to 15).

4.1.2 Results and discussions

Accordingly, the 68,000 ten-minute data were gathered for the categorization. The integrated D_v which is the summation of the luminance of 145 sky points was used to form the L_z/D_v ratio. Table 4.1 summarizes the number of normal and opposite for all 15 standard skies. After applying all the above tests, about 24,000 and 10,300 data points were classified as normal and opposite cases, respectively. This means that around 50% of the data (i.e. 33,700 out of 68,000) were found outside the ±2.5% zone and excluded from the analysis. The excluded data were mainly of clear sky conditions at high α_s. It is not surprising given that Hong Kong is located at low-latitude region (22.3°N) and high solar elevations of more than 35° appear frequently for most of daytime. It can be seen that the normal cases are more than the opposite ones under overcast and clear skies and vice-versa for partly cloudy conditions. Such features are quite similar to those for Bratislava in winter [Bartzokas et al., 2003]. The frequency of occurrence shows a variation ranging from 3.1% for sky type 15 to 10.3% for sky types 1 and 2. The RMSEs are quite large for individual skies (including normal and opposite cases) ranging between 24.5% for sky standard 13 and 51.3% for sky standard 9. The overall RMSE is 34.6% which is 10.8% higher than that based on the best-fitting approach.

Table 4.1 Frequency of occurrence for the 15 standard skies of Hong Kong using Bartzokas et al. method

Sky standard	No. of occurrence			Frequency of occurrence, %	RMSE, %
	Normal	Opposite	Total		
1	3107	412	3519	10.3	30.3
2	3083	434	3517	10.3	35.0
3	2709	439	3148	9.2	35.1
4	2760	469	3229	9.4	33.4
5	1631	511	2142	6.2	49.4
6	1848	561	2409	7.0	35.8
7	580	1664	2244	6.5	35.5
8	860	1305	2165	6.3	37.0
9	735	739	1474	4.3	51.3
10	985	764	1749	5.1	38.1
11	1495	723	2218	6.5	28.5
12	1342	748	2090	6.1	27.7
13	1240	713	1953	5.7	24.5
14	907	462	1369	4.0	27.6
15	672	402	1074	3.1	35.3
Total	23954	10346	34300	100	34.6

Figure 4.1 shows the detailed frequency distribution of the various sky types for every α_s. The maximum of frequency of occurrence appears in sky number 1 at α_s = 45-50^o in "normal" classification. A general examination of the figure reveals that most of the cases are identified as the overcast sky conditions (i.e. sky numbers 1 - 5), particularly when the α_s is lower than 55^o. For the partly cloudy (i.e. sky numbers 6 - 10) and clear sky (i.e. sky numbers 11 - 15), the most frequent cases occurs at α_s greater than 65^o and at α_s from 35^o to 50^o respectively. Under "opposite" classification, the peak occurrence appears at higher α_s, with sky types 2, 7, 8, 10 and 12 being predominant.

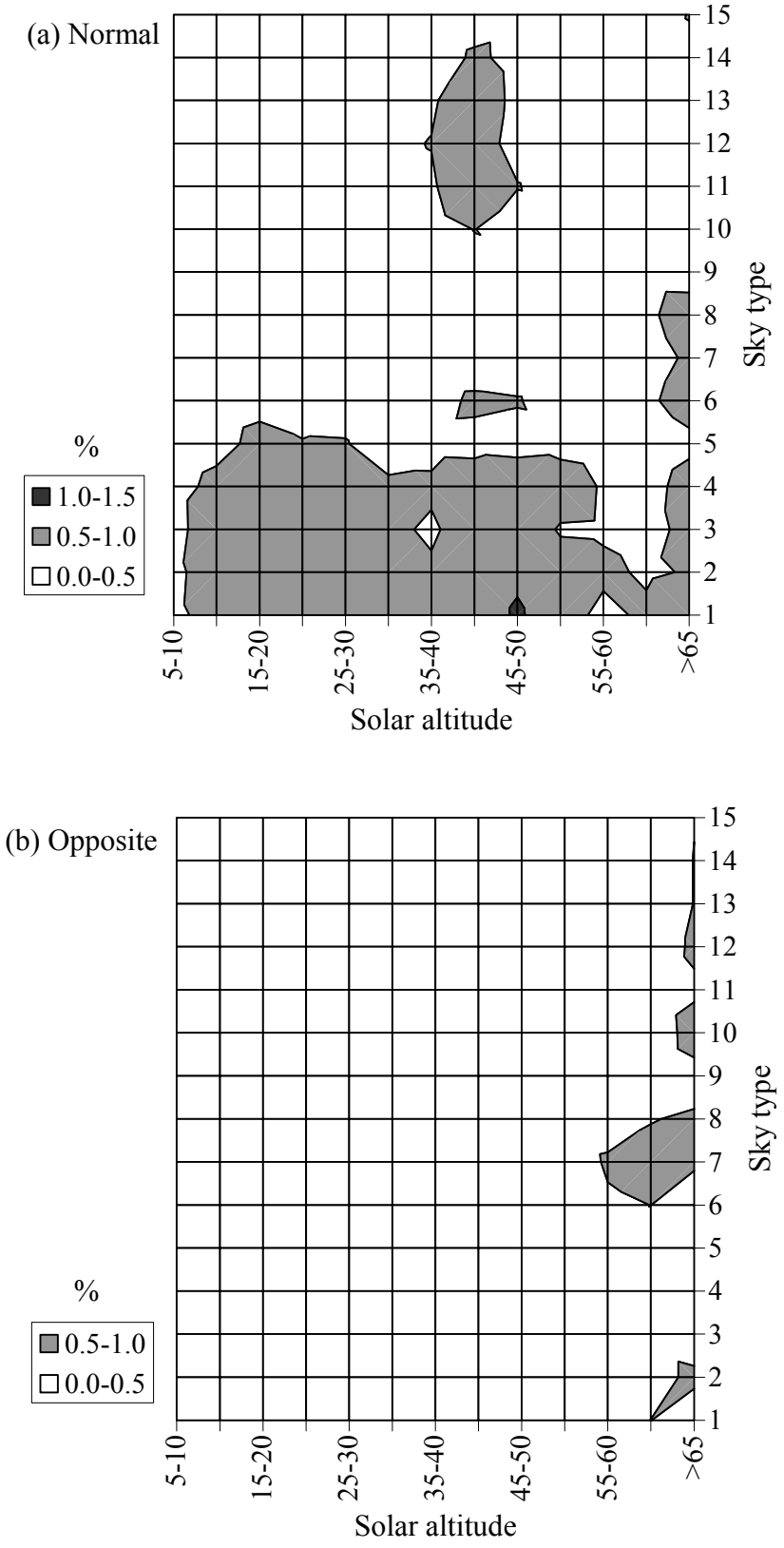

Figure 4.1 Frequency distribution for the 15 standard skies under (a) normal, (b) opposite classification

4.2 Proposed Approach (15 Standard Skies)

The Bartzokas et al. [2003; 2005] approach is appropriate for sky classification. However, for places of low-latitude climates, substantial data may be excluded for further analysis. As proposed by Kittler and Darula [2002], the pair L_z/D_v-G_v/E_v is an effective hybrid daylight-variable to interpret the 3 typical sky conditions. There are no clear-cut values for L_z/D_v-G_v/E_v to represent various sky conditions. Based on the measured Hong Kong data, therefore, Li and Lau [2007] proposed the L_z/D_v-G_v/E_v ranges for identification of the 3 typical skies. Table 4.2 summarizes the criteria adopted for the analysis.

In the following sections, a new approach based upon the L_z/D_v-G_v/E_v ranges proposed by Li and Lau [2007] are developed to further identify the 15 standard skies. Common climatic variables described in section 3.1 (i.e. α_s ,L_z/D_v, G_v/E_v, D_v/E_v and T_v) are adopted. A set of criteria for standard sky classification is suggested and the performance of the proposed approach is assessed by RMSE.

Table 4.2 Classification criteria for the 3 typical sky conditions using hybrid sky parameter L_z/D_v-G_v/E_v

Sky conditions	Proposed range
Overcast	$L_z/D_v \geq 0.3$ and $G_v/E_v \leq 0.3$
Partly cloudy	$0.17 < L_z/D_v < 0.3$ and $G_v/E_v \leq 0.3$ or $L_z/D_v > 0.17$ and $0.3 < G_v/E_v < 0.5$
Clear	$L_z/D_v \leq 0.17$ or $L_z/D_v \geq 0.17$ and $G_v/E_v \geq 0.5$

4.2.1 Methodology

The climatic data firstly can be categorized into 3 general sky conditions by applying the L_z/D_v-G_v/E_v ranges proposed by Li and Lau [2007]. To distinguish between the 5 sky types under the overcast database, the ratios of L_z/D_v and D_v/E_v were employed. Darula and Kittler [2004a] showed that the usual L_z/D_v ranges for sky standards 1 to 4 are greater than 0.32 and $L_z/D_v = 0.32$ represents absolute uniform sky (i.e. sky type 5) [Li et al., 2004a]. It indicates that L_z/D_v ratio of 0.32 would be an appropriate threshold to identify sky standard 5. Sky standards 1 to 4 can further be categorized by using D_v/E_v ratio. As indicated in the table of Standard Sky Luminance Distribution (SSLD) [Kittler et al., 1998] the usual D_v/E_v ratio spays is greater than 0.18 for sky standards 2 and 4, and less than 0.18 for sky standards 1 and 3.

Due to different gradation function as well as the same scattering indicatrix function of sky standards 1 and 3, the L_z/D_v ratio for sky standard 1 and 3 were normally over 0.38 and below 0.38, respectively, meaning that $L_z/D_v = 0.38$ would be the best indicator to separate those skies. A thorough study of Figure 3.1 indicated that the theoretical L_z/D_v curves can be used to separate sky types sky types 2 and 4. Although there is a slight brightness rising towards the solar position in sky standards 2 and 4, their theoretical L_z/D_v curves would not intersect each other in any α_s and hence the remaining climatic data can be recognized accordingly. Figure 4.2 presents the flow chart to identify the 5 overcast skies.

For partly cloudy conditions, sky standard 6 should be without any direct sunlight. When there is no direct component, it can be considered as sky standard 6. The remaining 4 sky standards can be classified by using T_v. The typical T_v values for sky

standards 7 and 9 are generally more than those for sky standards 8 and 10, and T_v of 12 is the criterion. Then, the theoretical L_z/D_v curves are employed again to interpret the sky standards. Figure 4.3 presents the proposed criteria for the classification of the partly cloudy skies.

Likewise, such daylight variables were also adopted to categorize the 5 clear sky standards. T_v of less than 4 is an appropriate indicator to single out sky standard 12 and sky type 15 can be differentiated when D_v/E_v ratio is 0.3 or more. Next, the theoretical L_z/D_v curves are employed to pick out sky standard 14. As the usual T_v ranges from 2.5 to 6.5 for sky standard 11 and between 3 and 8 for sky standard 13 [Kittler et. al, 1998], ultimately, T_v of 5.5 is applied again to separate the sky standards 11 and 13. The approach to classify the clear sky standards is summarized in Figure 4.4.

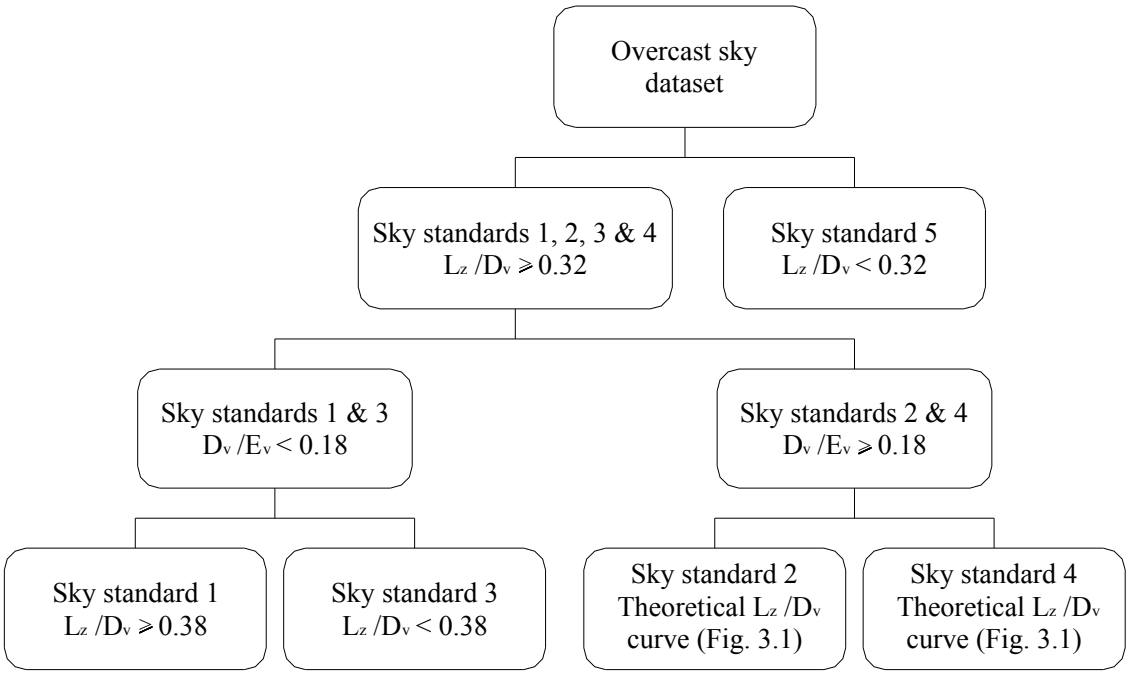

Figure 4.2 The proposed classification criteria for the 5 overcast skies

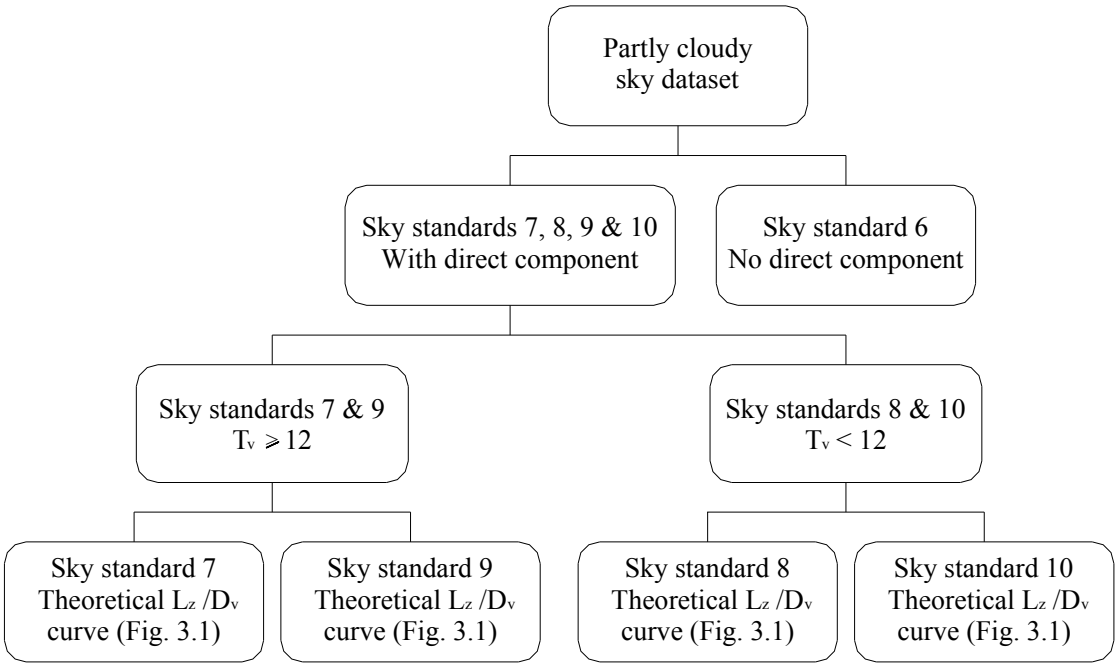

Figure 4.3 The proposed classification criteria for the 5 partly cloudy skies

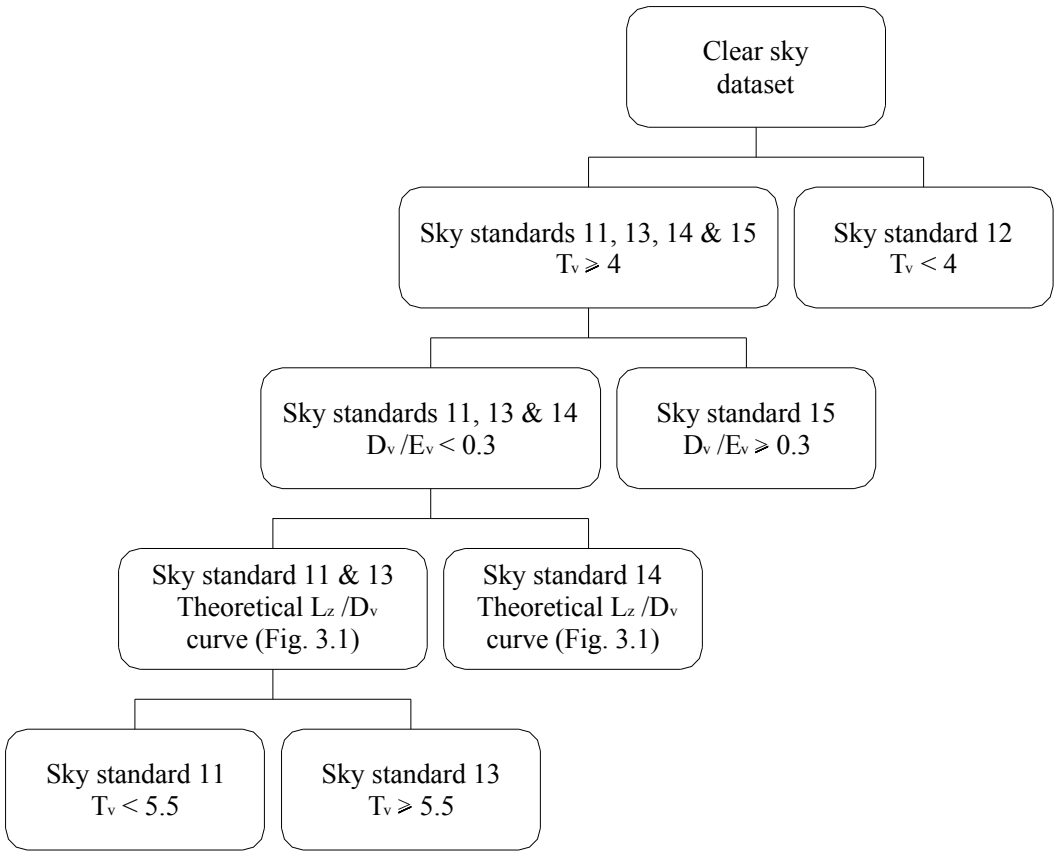

Figure 4.4 The proposed classification criteria for the 5 clear skies

4.2.2 Results and discussions

The same 68,000 ten-minute sky luminance data were used to form and evaluate the sky classification in accordance with the proposed methods. Again, the RMSEs were computed for an assessment. Figure 4.5 displays (a) the frequency of occurrence and (b) the RMSE results. Sky standards 1 and 13 have the largest frequency of occurrence representing over 33% of all the skies considered. The findings are in good agreements with our previous works [Li et al., 2004a; Li et al., 2004b]. The sky standard 10 has the lowest frequency of occurrence of 1.4%. Apart from sky standards 8 and 9, the frequency of occurrence for individual sky is similar to that using best-fitting method (i.e. Figure 3.6).

Generally, Figure 4.5 indicates that the sky standards are quite evenly distributed in terms of the 3 typical sky conditions under such categorizations. The RMSEs range from less than 25% for sky standard 14 to just over 49% for sky standard 9. The overall RMSE is 32.4%, representing 2.2% improvement compared with the approach suggested by Bartzokas et al. [2003; 2005] and no data are further rejected using the proposed approach.

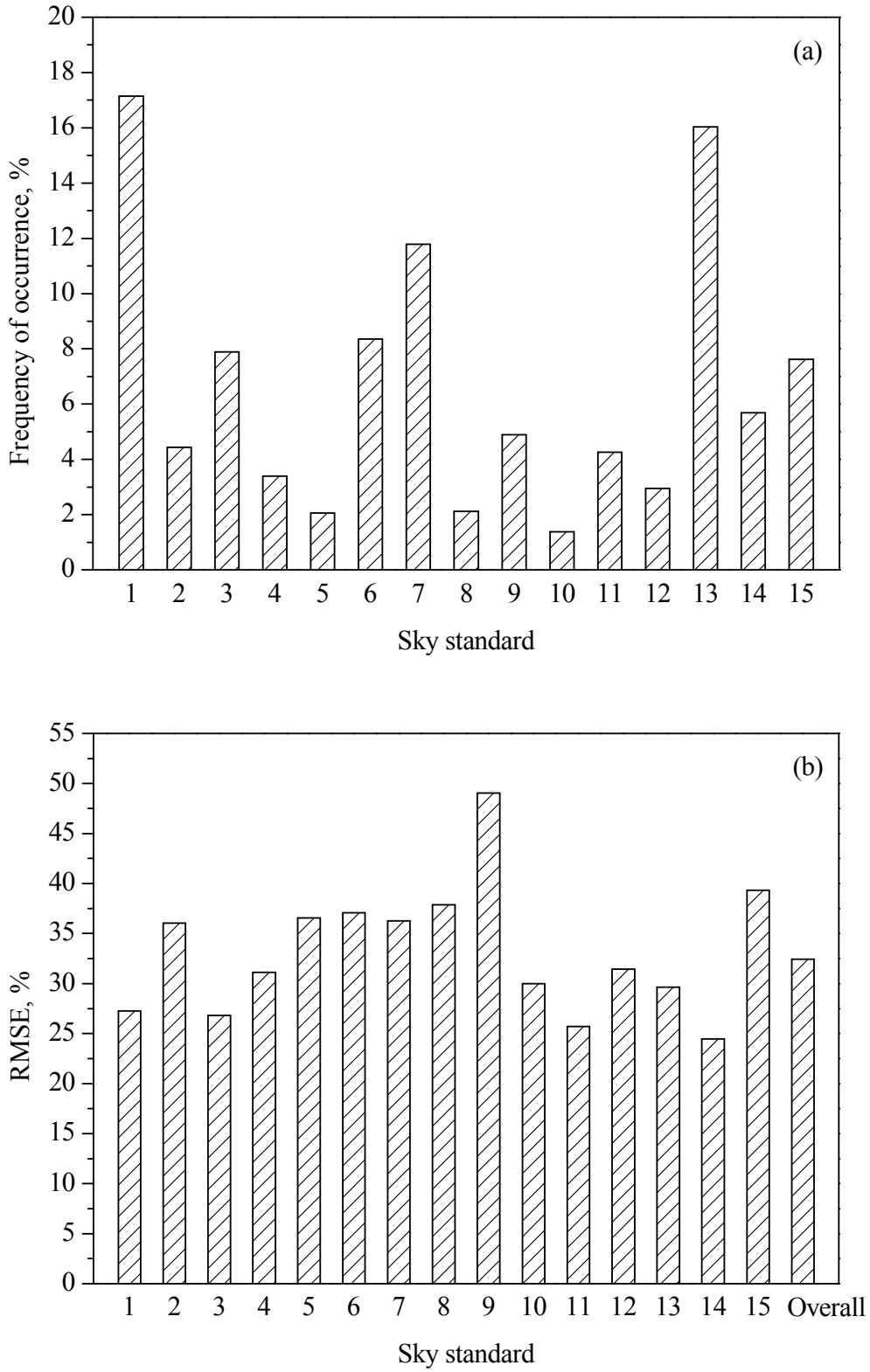

Figure 4.5 (a) Frequency of occurrence for the 15 standard skies, (b) RMSE of the luminance distribution of actual skies to the 15 standard skies using climatic parameters approach

4.3 Proposed Approach (6 Prevailing Standard Skies)

As discussed in section 3.3.2, a subset of the CIE standard skies including overcast, partly cloudy and clear sky conditions are sufficient to describe the daylight climates at any particular site. Individual standard skies with high frequency of occurrence and low RMSE are appropriate to represent the prevailing sky conditions for a given location. As a result, sky numbers 1, 3, 6, 7, 8 and 13 were selected to represent the prevailing sky standards in Hong Kong. This section presents the work on the classification of prevailing sky luminance distributions for Hong Kong using the climatic parameters shown in section 3.1 excluding D_v/E_v ratio. A comparative assessment of the best-fitting and the proposed methods against sky luminance data is reported.

4.3.1 Methodology

Similar with the approach recommended in the preceding section, the climatic data are first split into 3 general categories by following the criteria proposed by Li and Lau [2007]. To distinguish between sky numbers 1 and 3 under the overcast database, the ratio L_z/D_v was employed again. Darula and Kittler [2004a] showed that the usual L_z/D_v spay is greater than 0.38 for sky standard 1 and 0.33 to 0.38 for sky standard 3. It indicates that a value of the L_z/D_v of 0.38 is the appropriate threshold. For partly cloudy conditions, sky standard 6 should be without direct-beam daylight illuminance. When there is no direct-beam component, it can be considered as sky standard 6. The remaining two partly cloudy sky standards (sky numbers 7 and 8) can be classified by using T_v. The typical T_v values range between 5 and 12 for sky number 8 and around 12 for sky number 7. The clear-cut T_v value is 12. The proposed criteria of L_z/D_v, P_v and T_v for the classification of the standard skies 1, 3, 6, 7 and 8 are summarized in Figure 4.6.

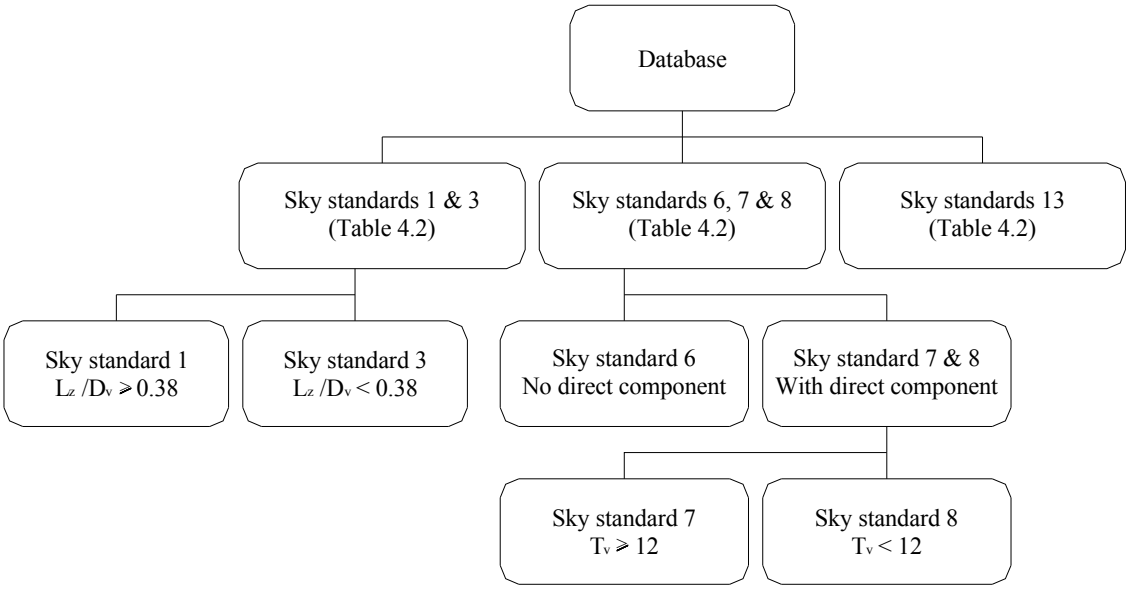

Figure 4.6 The proposed classification criteria for the prevailing set of 6 standard skies

4.3.2 Results and discussions

The same 68,000 ten-minute sky luminance data were used to form and evaluate the classification of the 6 prevailing sky standards (sky numbers 1, 3, 6, 7, 8 and 13) in Hong Kong. Again, the RMSEs were computed for an assessment. Figure 4.7 displays (a) the frequency of occurrence and (b) the RMSE results. Comparing the results to the findings of best-fitting approach, it is seen that they are similar. The most frequent sky type is 13, it obtains the largest frequency of occurrence of 36.5%. For sky standards 1, 3 and 6, the frequency of occurrence varies slightly. Overcast (i.e. sky numbers 1 and 3) and partly cloudy skies (i.e. sky numbers 6, 7 and 8) contribute about 35% and 29% of the Hong Kong sky conditions, respectively. The sky standards are quite evenly distributed in terms of the 3 typical sky types under such categorizations However, the main difference is that the least frequent sky type determined by the proposed approach is sky number 8 instead of sky number 7 given by best-fitting approach. The frequency

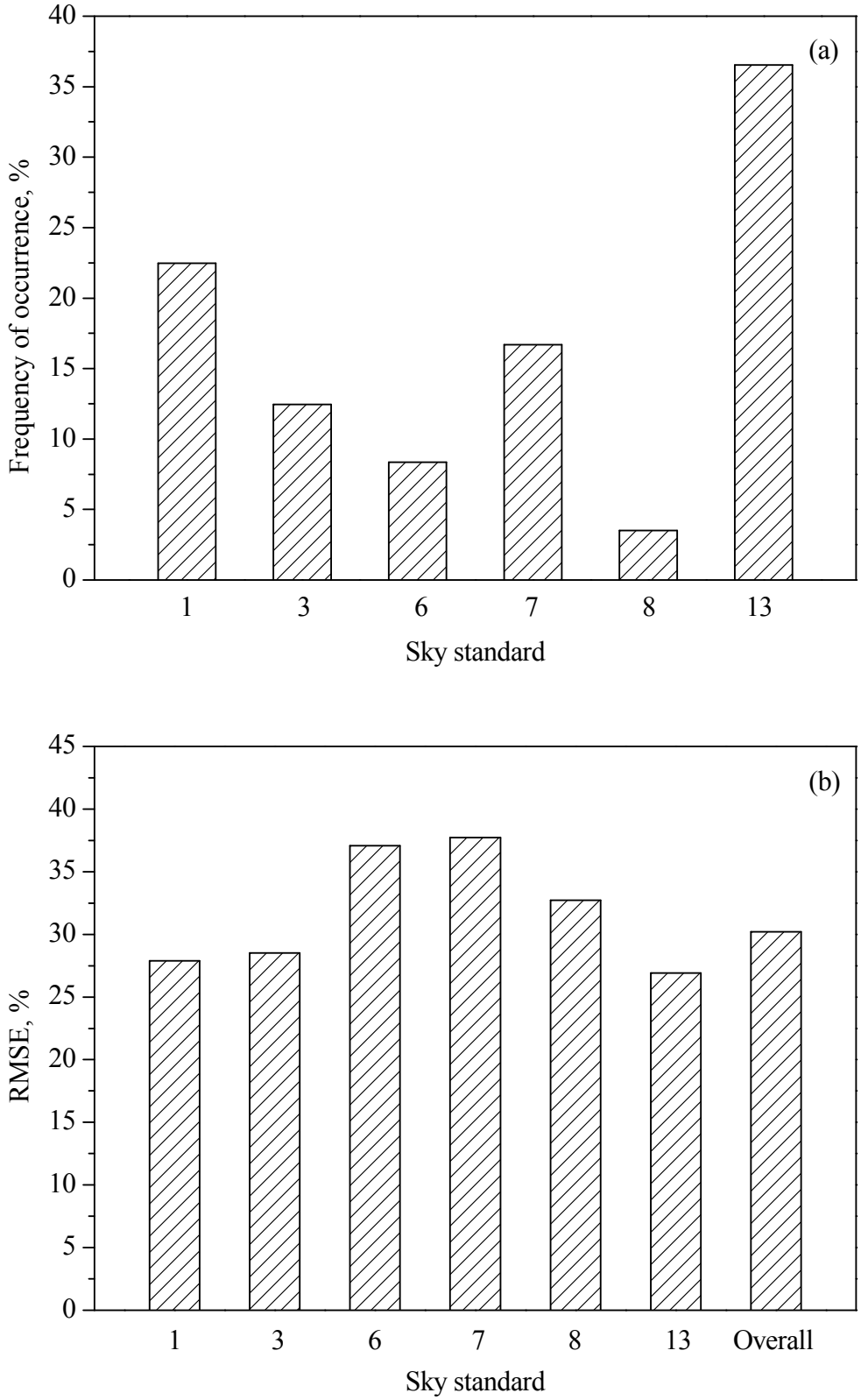

Figure 4.7 Frequency of occurrence for the prevailing set of standard skies, (b) RMSE of the luminance distribution of actual skies to the prevailing set of standard skies using climatic parameters approach

of occurrence for sky number 8 drops rapidly from 10.8 to 3.5% while sky number 7 rises from 9.5 to over 16.7%. The RMSEs range from 26.9% for sky standard 13 to 37.7% for sky standard 7. A general inspection of the figure reveals that the RMSEs of all the partly cloudy skies are over 32%. The overall RMSE for the reduced set of the 6 prevailing sky standards in Hong Kong is 30% which is 5% greater than the finding determined by the best-fitting approach.

4.4 Summary

The performance of the proposed approach and the Bartzokas et al. method was evaluated against the measured data in terms of RMSE. The statistical analysis revealed that using the Bartzokas et al. method, around 50% of the data were rejected for sky classification and the overall RMSE was calculated of around 35%. Furthermore, a series of range of climatic parameters have been suggested to recognize the sky conditions in Hong Kong. The proposed classification criteria were relied on the features and limitations of the parameters. The proposed approach for the complete set of 15 standard skies produced an overall RMSE of just less than 33% without data reduction whereas the RMSE for the prevailing set of 6 standard skies was 30%. The approach recommended in this chapter provided a clear and straightforward procedures to recognize the standard sky types. For pattern classification, however, some researchers suggested a computational tool, artificial neural networks which have widely been used in many classification problems. In chapter 5, the usefulness of ANNs as a tool for the sky classification will be examined and the performance of ANN will be compared with the results obtained in this chapter.

Chapter 5

SKY CLASSIFICATION II:

ARTIFICIAL NEURAL NETWORKS

ANNs have emerged as a useful tool for classification, modeling, pattern recognition, multivariate function approximation and forecasting. According to Haykin [1999], the neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge, making it available for use. It consists of a large number of processing elements called neurons which are interconnected in a layered parallel structure to build a network. The network attempts to imitate the characteristics of the human brain and nerve system which learn from experience [Lim et al., 2003]. By providing both of the input and output exemplars, the network learns by processing the input exemplar and comparing the result with the desired response. In case of any differences between the outcome and expectation occur, the network architecture and the weights of each interconnected link are modified to obtain the optimal performance.

In most applications, ANNs can handle large and complex data sets with many interrelated parameters running much faster than dynamic simulation programs and can

give more accurate results as compared with the traditional statistical models [Masters, 1994]. Apart from the outstanding performances, there is no constraint on the number of input variables and ANNs can work with noisy, overlapping, highly nonlinear and non-continuous data [Zaknich, 2003].

Neural networks are of interest to researchers as well as professional parties in many fields of study including industrial engineering [Quteishat and Lim, 2008; Nabhani and Shaw, 2002], climatology studies [Cam et al., 2005; Nasseri et al., 2008] and medicine [Hayashi et al., 2000; Guler and Ubeyli, 2007], etc.

Instead of using a single type of neural network for different purposes, some researchers prefer to integrate two or more neural networks to form a hybrid system which can alleviate the drawbacks or limitations of each network without losing their own properties and merits. Lim and Harrison [1997] proposed a hybrid network, called Probabilistic Fuzzy ARTMAP (PFAM) constructed by the Fuzzy ARTMAP neural network (FAM) and PNNs, for the aims of on-line learning and probability estimation tasks. In both on-line and off-line learning, the results indicated that PFAM performed better and it showed a significant improvement when compared with the single FAM.

Lee et al. [2006] developed a GRNNFA model which combines the Fuzzy adaptive resonance theory (FA) and the General regression neural network model (GRNN). The GRNNFA model was applied to predict the occurrences of flashover when the dimensions of the compartment and maximum heat release rate were served as the inputs of the model. Based on the 375 data samples, the authors reported that the prediction results were excellence.

In preceding chapter, the 15 CIE standard skies were classified using various appropriate climatic parameters but the criteria to distinguish individual skies are not always clear-cut and may lead to ambiguous findings. The aim of this chapter is to evaluate the performance of PNNs on identification of standard sky luminance distributions. The network architecture and applications of PNNs were introduced. Apposite daylight variables including α_s, zenith luminance, horizontal global and diffuse illuminance, and luminous turbidity were employed for the sky categorization and a study of parameter analysis was presented. The potential of PNNs for standard sky classification was investigated.

5.1 Probabilistic Neural Networks

In fact, there are two types of neural networks which are commonly used for pattern classification in the literature. The first one is multilayer feed-forward neural networks with back-propagation (BP) learning algorithm and the second one is PNNs. Back-propagation neural networks usually employ heuristic approaches to determine underlying class statistics, and those approaches normally consists of many small incremental modifications to network parameters so as to improve the model performance gradually [Specht, 1990b]. Although it is a well-known and popular supervised neural network in many applications, numerous researchers argued that BP neural network is a "black box" that the classification principles may be difficult to justify [Calderon and Cheh, 2002]. The inability to explain how and why the conclusions are obtained is one of the restrictions to apply BP neural networks [Salchenberger et al., 1992].

Instead of neural networks with BP, many researchers are preferred to use another tool, PNNs, for the purpose of pattern classification [Gaganis et al., 2007]. PNN is based on well developed statistical principles rather than heuristic approaches. It implements the Bayesian decision strategy for classification and nonparametric estimator to estimate the probability density function (PDF). Also, the training process of PNNs is one-pass and without any iteration for weight adaptation.

When compared with the conventional BP, the processing speed of PNN has been proven to be faster [Specht and Shapiro, 1990] while the ability of the networks to generalize an unknown pattern is nearly the same [Specht and Shapiro, 1991]. The training and testing processes of the PNN are easier because of the simplified architecture and only one parameter is involved for optimalizing network performance. The main disadvantage of PNNs are the computational memory consumption as well as the processing speed is directly proportional to the size of the training set. Since all the training vectors are copied to the pattern neurons during the training phase, the more training vectors are involved, the higher the requirements of the computer. Nevertheless, with most average computer speed available today, the processing time can be minimized effectively [Hajmeer and Basheer, 2002]. As a new user to apply the neural networks and after considering the objectives of the study, PNN is selected as a first step to determine the potential of neural networks for sky classification. In fact, PNN has been widely used as a pattern classifier in various disciplines.

In ocean engineering, Kim et al. [2008] estimated the stability number of armor blocks of breakwaters using PNNs. The study was based on the 5 parameters, namely the notional permeability of the breakwater, the damage level, the surf similarity parameter,

the water depth and the spectral shape, prepared by Van der Meer [1988]. The performance of PNNs was compared with the results obtained from empirical models and the other types of neural networks. The author concluded that the predictive capability of PNN was the best.

In microbiological applications, Hajmeer and Basheer [2002] demonstrated the potential of applying PNNs for categorizing the bacterial growth and no-growth data. Using the 179 data samples including 99 cases of growth and 80 cases of no-growth, the PNNs were trained under 4 training schemes and all of them yielded highly accurate PNNs with the percentage correct not less than 94%.

5.1.1 Overview of PNNs

PNNs is a feed-forward neural network developed by Specht [1990a]. It is a model based on Bayesian decision strategy [Mood and Graybill, 1962] for decision making and Parzen's theorem [Parzen, 1962] for estimating the PDF. For the two classes A and B, the Bayesian's strategy identifies an unknown sample into population class A if

$$h_A c_A f_A(\mathbf{x}) \quad > \quad h_B c_B f_B(\mathbf{x}) \tag{5.1}$$

where h_A = prior probability of the sample belongs to class A

h_B = prior probability of the sample belongs to class B

c_A = cost of misclassification for class A

c_B = cost of misclassification for class B

$f_A(\mathbf{x})$ = probability density function for class A

$f_B(\mathbf{x})$ = probability density function for class B

The prior probabilities and the cost of misclassification are usually treated as the same for each class which contains the same number of data samples. Therefore, estimation of the PDF for each class is the last task for implementing the Bayesian decision strategy. However, the PDF, which defines the boundaries for each data class, is not always known. Parzen PDF estimator is a useful tool to estimate PDF from the data samples but it can only be used to determine univariate functions. In multivariate cases, Cacollous [1966] extended Parzen PDF estimator and suggested that the PDF can be expressed as equation 5.2 if Gaussian kernel function is used.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\sigma^{p} n} \sum_{j=1}^{n} \exp\left[-\frac{(\mathbf{x} - \mathbf{x}_{Aj})^{T}(\mathbf{x} - \mathbf{x}_{Aj})}{2\sigma^{2}} \right] \tag{5.2}$$

where n = total number of training pattern

σ = smoothing factor

p = dimension of input vectors

$\mathbf{x}_{Aj}$ = j^{th} training vector from class A

Instead of sigmoid function, Gaussian function is used and assigned at each pattern neuron in the training data samples. The PDF for each class can be computed by summing the Gaussian function belonging to the same class. The smoothing parameter (or kernel width) σ, estimated from the validation set or simply by trial and error, is the most important element to determine the width of the activation function [Chen and Hsu, 2007]. Different values of the smoothing parameter may cause different impacts on the degree of interpolation that occur between adjacent pattern vectors. The influence of the choice of smoothing parameter has been examined by Specht [1990a].

5.1.2 The PNNs architecture and operation

Figure 5.1 presents the basic architecture of the PNNs that will be used in the study discussed in next section. It consists of an input layer, a pattern layer, a summation layer and an output layer. The number of neurons in the input layer depends on the total number of the variables needed to depict the samples to be classified. The input and the pattern layer are fully interconnected so that the *p*-dimensional training vectors can be distributed to the pattern layer one by one.

For the pattern layer, the number of neurons is equal to the number of training vectors. The pattern neurons are used to store the training vectors, and the weight for each pattern neuron is the element of the corresponding training vector. The summation layer contains one summation neuron for each class, and each neuron is used to collect and sum the outcomes from all pattern neurons belonging to same category. The output of the summation neuron for a particular class is actually the estimated PDF for that class. The output layer contains only one neuron where the unknown input vector are classified to one of the classes that has the highest probability density function in the summation layer using the Bayesian decision strategy.

Unlike other types of neural networks, the training stage of the PNN actually just copies the training vectors to the pattern neurons and determines the optimal value of the smoothing factor. Once a new test vector, $\mathbf{x}$, is presented to the PNNs for classification, the ith pattern neuron forms a dot product $\mathbf{Z}_i$ of the $\mathbf{x}$ with a weight vector $\mathbf{w}_i$.

$$\mathbf{Z}_i = \mathbf{x} \cdot \mathbf{w}_i \tag{5.3}$$

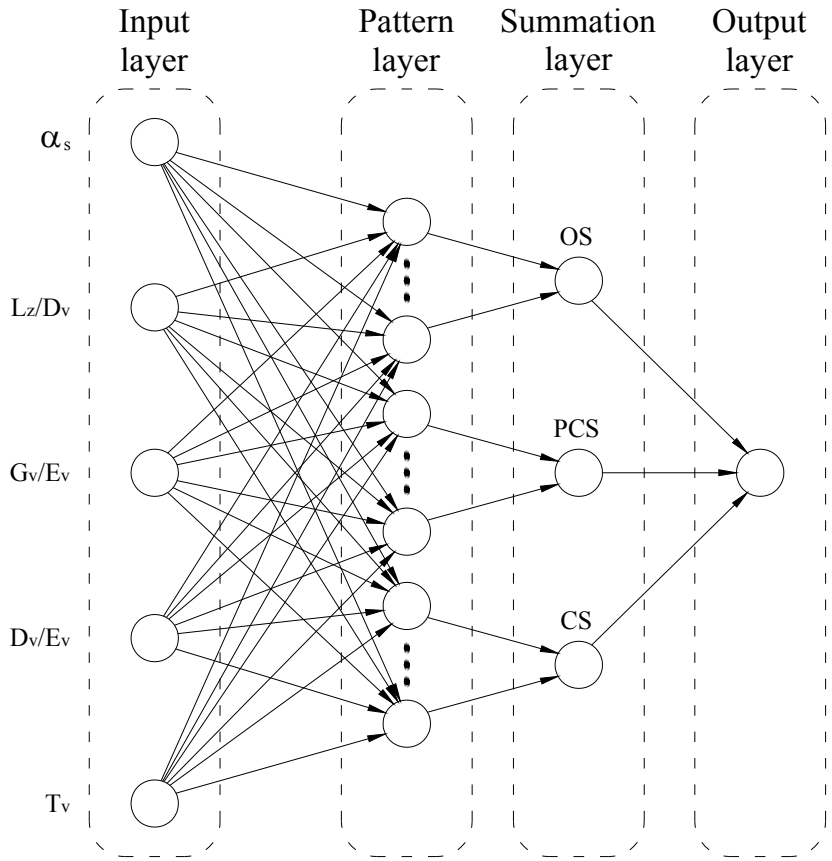

Figure 5.1 Basic configurations of PNNs
(OS – overcast sky, PCS – partly cloudy sky and CS – clear sky)

Then $\mathbf{Z}_i$ processes through a nonlinear transfer function as shown in Figure 5.2.

$$g(\mathbf{Z}_i) = \exp\left(\frac{\mathbf{Z}_i - 1}{\sigma^2}\right) \tag{5.4}$$

If $\mathbf{x}$ and $\mathbf{w}_i$ are normalized to unit length,

$$\exp\left(\frac{Z_i - 1}{\sigma^2}\right) = \exp\left[-\frac{(\mathbf{x}_i - \mathbf{w}_{i,j})^{\mathrm{T}}(\mathbf{x}_i - \mathbf{w}_{i,j})}{2\sigma^2}\right] \tag{5.5}$$

which is the same form as the exponential function of equation 5.2.

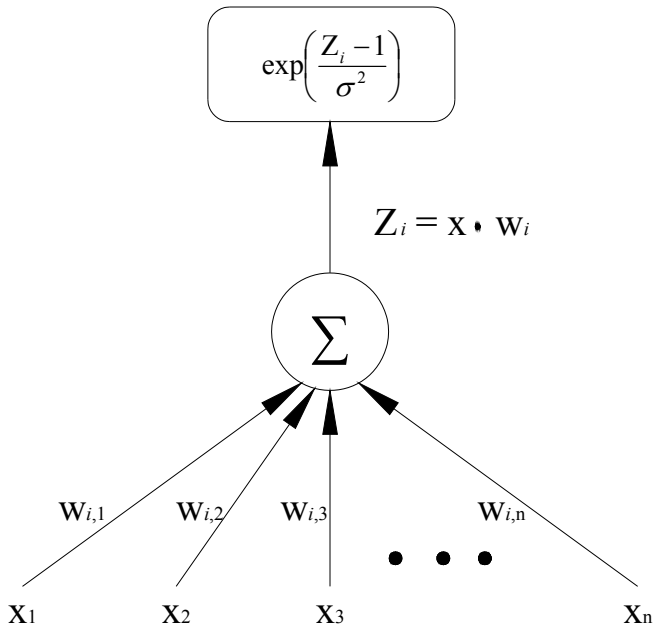

Figure 5.2 Internal structure of pattern neuron

It should be noted that the denominator of the exponential function (Equation 5.5) denotes the Euclidean distance between the test vector and weight vector stored in the pattern layer.

5.1.3 Methodology

Daylight illuminance and sky luminance data measured during the 84-month period from January 1999 to December 2005 were used to develop and evaluate the PNN models for sky classification. In reality, the occurrences of the standard skies were not evenly distributed as shown in Figure 3.6. Sky number 1 contributed about 18% of the Hong Kong sky conditions whereas just over 1% for sky number 9. However, Swingler [1996] claimed that the training set should contain nearly the same number of training vectors from each class. To avoid bias in the over represented classes, only 600 data

samples for each identified sky standard were randomly selected from the 68,000 database. By doing so, the extra data samples in the over represented classes were removed to ensure that the databases are not bias [Basheer and Hajmeer, 2000]. In total, 9,000 data samples were gathered and used in this study. Five climatic parameters as described in section 3.1 were used as the input parameters of the PNNs to discriminate the 15 standard skies. To make sure that each input parameter within the same range, all the data samples were scaled linearly within the range from zero to one during the data preprocessing phase.

To develop PNNs as a classifier for recognizing different sky types, the database was firstly divided into 3 subsets, namely training set, validation set and test set. Although the dimension of each set has significant impacts on the classification accuracy, there are no practical rules or theories to determine the size of each set, and just only some empirical rules of thumb are exist [Basheer and Hajmeer, 2000]. Large validation set may improve the generalization capability of the network but the remaining small training set may not provide sufficient information to train the network and vice versa. To balance the effects, 4,500 and 1,800 out of 9,000 data samples (i.e. 50% and 20% of the database) were randomly extracted to represent the training and validation set respectively. Both the training and validation set were involved in the training of the PNN. The remaining 2,700 data samples were the test set which was never used in PNN development. The PDF for each class were defined by all the training vectors in the training set while the optimal value of the smoothing factor were estimated by the validation set. For the present study, a software package “NeuroShell 2” [NeuroShell 2, 2000] is used to train and develop the neural networks for sky classification. To determine the optimal value of the smoothing parameter for the PNNs, the software

recommends genetic algorithm (GA) to improve the generalization capability of the network even though the training duration will be much longer. A population of individuals is firstly generated by randomly selecting different genes. Then, propagation techniques are applied to the individual with the highest fitness so as to generate a new population. The above steps are repeated until an optimal smoothing parameter is obtained. The details about the smoothing parameter selection using GA can be referred to Mao et al. [2000].

Hansen et al. [1992] and Kalatzis et al. [2005] reported that the accuracy as well as generalization capability of neural network could be improved significantly by integrating a number of models. For classification problems, majority voting and weighted voting are two commonly used ensemble approaches [Osowski et al., 2008]. The majority voting scheme was employed in this study since each PNN model has the same influence on the ultimate classification results. The basic idea of majority voting is to train a set of models and allow them to vote [Ravi et al., 2008]. To undertake the majority voting scheme, 30 different pairs of training and validation sets were prepared to train the PNNs and hence 30 trained PNN models were built. For the purpose of clarity, PNN1 represents the PNN model with the combination of first training and validation set, PNN2 represents the PNN model with the combination of second training and validation set, PNN3 represents the PNN model with the combination of third training and validation set and so on. All the 30 trained PNN models were tested independently by the same test set. The results collected from each of the PNN models were combined to form an ensemble system and the ultimate predicted class of a particular test data was obtained after majority voting. Figure 5.3 shows the overview of PNNs based on majority voting scheme.

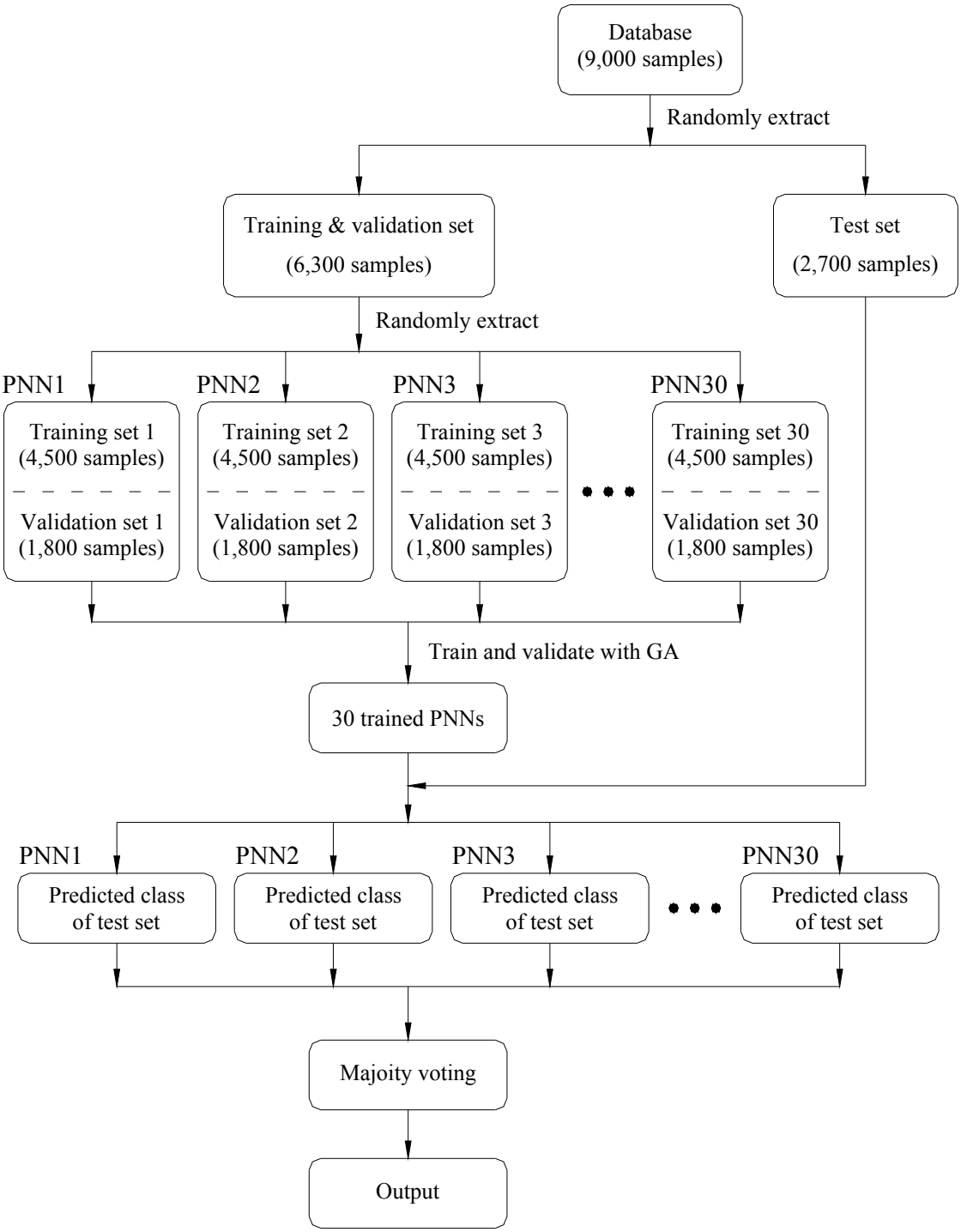

Figure 5.3 Probabilistic neural network with majority voting

5.2 Results and Discussions

In the present study, PNN models trained by 30 pairs of training and validation sets were employed firstly to categorize the 3 typical sky conditions. The results of the PNN models were then integrated by applying the majority voting scheme. Afterwards, the investigations were extended to the 15 CIE standard skies. For each analysis, a confusion matrix was constructed to evaluate the performance of the models. The number of rows and columns of the matrix are equal to the number of classes in each study. The rows and columns of the table actually represent the actual sky type and the predicted sky type respectively. The numbers of correct classification appear only in the diagonal positions while the numbers of misclassification indicate in the off-diagonal positions. From the confusion matrix, the hit percentage, denoted as the ratio of the number of samples correctly classified to the total number of samples to be classified, can be computed accordingly.

5.2.1 3 typical sky conditions

Table 5.1 shows sky classification results over the 30 PNN models. It can be seen that the variations of the successful rate among the 30 PNN models for the training, validation and test sets are quite small. The hit percentage lies between just below 91% and 96% using the training and validation sets, and from just over 87% to 89% based on the test set. As expected, the classification results of the training and validation sets were always better than the corresponding test set.

To further evaluate the performance of PNNs, the classification results based on 2,700 test data under the 30 PNN models were gathered and the results are presented in the

confusion matrix shown in Table 5.2. The sky classification interpreted by PNN models are in good agreement with those using best-fitting approach. For instance, 856 out of 899 clear skies are correctly classified. Only 1 and 42 skies were identified as overcast and partly cloudy conditions respectively. Similar findings were obtained for overcast skies. For the partly cloudy skies, 88 and 76 out of 908 cases were respectively categorized as overcast and clear skies. In general, the PNN models show the best performance for distinguishing clear skies, then overcast skies, and finally partly cloudy skies. The hit percentage of the PNN models is 89.2%.

Table 5.1 Classification results of the 30 PNN models for the 3 typical sky conditions

	Hit percentage, %			Hit percentage, %	
PNN	Training & validation set	Test set	PNN	Training & validation set	Test set
1	94.8	88.4	16	94.6	88.3
2	94.7	88.9	17	90.7	87.9
3	93.9	88.7	18	94.8	88.5
4	95.1	88.9	19	94.4	88.9
5	91.1	87.1	20	95.0	88.4
6	94.6	88.3	21	95.0	89.1
7	93.5	88.1	22	95.3	88.7
8	95.1	89.1	23	94.5	88.2
9	91.7	88.2	24	95.3	88.6
10	94.7	89.1	25	95.6	88.7
11	95.2	88.1	26	94.7	88.5
12	94.1	88.5	27	91.4	87.9
13	95.9	88.6	28	91.3	88.1
14	95.5	88.6	29	91.7	88.2
15	94.3	88.8	30	93.7	88.5

Table 5.2 Confusion matrix of the PNNs for the 3 typical sky conditions based on the test set (percentage of each sky type)

Actual sky type	Predicted sky type		
	Overcast sky	Partly cloudy sky	Clear sky
Overcast sky	809 (90.6%)	83 (9.3%)	1 (0.1%)
Partly cloudy sky	88 (9.7%)	744 (81.9%)	76 (8.4%)
Clear sky	1 (0.1%)	42 (4.7%)	856 (95.2%)

5.2.2 15 CIE standard skies

Likewise, the same 9,000 samples were used for the classification of the 15 CIE standard skies. In total, 15 output nodes were set. Again, 30 PNN models were carried out and Table 5.3 presents the findings. As expected, the successful rates were less than those for identified the 3 typical skies. The variation of the accuracy ranges between 87.6 and 89.1% for training and validation sets and from 64.3 to 67.1% for test set. Again, classification results based on the 30 PNN models for the 15 CIE standard skies were got together to examine the performance of PNN techniques and Table 5.4 illustrates the number of the test data successfully interpreted as individual sky types. The PNN models can effectively single out sky standard 11 with the correct classification percentage of 86.7% while the lowest correct classification percentage of 49.1% was found for sky standard 6. In total, 1,911 out of 2,700 data samples were correctly classified. This represents 70.8% of the hit percentage which is 18.4% lower than the classification of the 3 typical sky conditions.

Subsequently, the frequency of occurrence for the 15 standard skies, and the corresponding RMSE of the luminance distribution of actual skies to the 15 standard skies using the best-fitting and PNN approaches based on 2,700 test data were determined. Figure 5.4 demonstrates the findings. The frequency of occurrence ranges between 5.6 % for sky standard 9 and 7.6% for sky standard 7 via the best-fitting model and varies from 4.8% for sky standard 9 to 7.8% for sky standard 11 with PNN techniques. As expected, the RMSE values adopting PNN are generally more than those based on best-fitting technique. The peak difference of 8.3% appears in sky standard 5. The RMSE for all the 15 standard skies by best-fitting and PNN approaches are 24.9 and 28.2% respectively.

Table 5.3 Classification results of the 30 PNN models for the 15 standard skies

	Hit percentage, %			Hit percentage, %	
PNN	Training & validation set	Test set	PNN	Training & validation set	Test set
1	87.9	65.7	16	88.6	65.7
2	87.9	65.6	17	87.6	64.3
3	89.1	65.9	18	87.9	64.7
4	88.4	67.1	19	88.8	65.3
5	87.9	64.7	20	87.8	65.6
6	87.9	65.7	21	87.8	66.0
7	87.7	65.6	22	87.9	65.7
8	89.1	65.4	23	87.9	65.3
9	88.4	64.7	24	88.9	65.6
10	87.9	65.7	25	88.4	66.3
11	88.0	65.4	26	89.0	65.6
12	87.8	65.6	27	87.7	64.9
13	87.7	66.1	28	88.2	65.6
14	88.9	65.9	29	88.0	65.4
15	87.7	65.9	30	87.8	66.6

Table 5.4 Confusion matrix of the PNNs for the 15 typical sky conditions based on the test set (percentage of each sky type)

Actual sky type	Predicted sky type														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	127 (76.0%)	9 (5.4%)	14 (8.4%)	2 (1.2%)	12 (7.2%)	2 (1.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.6%)
2	13 (6.7%)	126 (65.3%)	10 (5.2%)	17 (8.8%)	14 (7.3%)	9 (4.7%)	0 (0.0%)	0 (0.0%)	1 (0.5%)	1 (0.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (1.0%)
3	18 (10.3%)	5 (2.9%)	133 (76.0%)	9 (5.1%)	7 (4.0%)	1 (0.6%)	2 (1.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
4	5 (2.8%)	7 (3.9%)	8 (4.4%)	116 (64.4%)	17 (9.4%)	15 (8.3%)	4 (2.2%)	2 (1.1%)	2 (1.1%)	3 (1.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.6%)
5	8 (4.5%)	8 (4.5%)	11 (6.2%)	12 (6.7%)	114 (64.0%)	12 (6.7%)	2 (1.1%)	1 (0.6%)	4 (2.2%)	3 (1.7%)	0 (0.0%)	1 (0.6%)	0 (0.0%)	0 (0.0%)	2 (1.1%)
6	2 (1.1%)	11 (6.3%)	4 (2.3%)	9 (5.1%)	16 (9.1%)	86 (49.1%)	22 (12.6%)	8 (4.6%)	5 (2.9%)	10 (5.7%)	1 (0.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.6%)
7	0 (0.0%)	2 (1.0%)	0 (0.0%)	8 (3.9%)	11 (5.4%)	15 (7.4%)	123 (60.3%)	13 (6.4%)	13 (6.4%)	11 (5.4%)	4 (2.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (2.0%)
8	0 (0.0%)	1 (0.5%)	0 (0.0%)	4 (2.1%)	2 (1.0%)	6 (3.1%)	15 (7.9%)	144 (75.4%)	3 (1.6%)	6 (3.1%)	6 (3.1%)	0 (0.0%)	3 (1.6%)	1 (0.5%)	0 (0.0%)
9	1 (0.7%)	2 (1.3%)	0 (0.0%)	3 (2.0%)	4 (2.6%)	15 (9.9%)	4 (2.6%)	8 (5.3%)	76 (50.0%)	17 (11.2%)	5 (3.3%)	6 (3.9%)	1 (0.7%)	2 (1.3%)	8 (5.3%)
10	0 (0.0%)	0 (0.0%)	3 (1.6%)	2 (1.1%)	4 (2.2%)	7 (3.8%)	13 (7.0%)	8 (4.3%)	14 (7.5%)	105 (56.5%)	11 (5.9%)	2 (1.1%)	4 (2.2%)	1 (0.5%)	12 (6.5%)
11	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (1.0%)	3 (1.5%)	3 (1.5%)	3 (1.5%)	5 (2.6%)	169 (86.7%)	6 (3.1%)	2 (1.0%)	1 (0.5%)	1 (0.5%)
12	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.6%)	2 (1.1%)	3 (1.7%)	2 (1.1%)	151 (84.8%)	11 (6.2%)	7 (3.9%)	1 (0.6%)
13	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (1.1%)	0 (0.0%)	2 (1.1%)	13 (7.1%)	9 (4.9%)	150 (82.0%)	5 (2.7%)	2 (1.1%)
14	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	7 (4.1%)	6 (3.5%)	148 (86.5%)	10 (5.8%)
15	1 (0.6%)	0 (0.0%)	1 (0.6%)	0 (0.0%)	0 (0.0%)	2 (1.2%)	0 (0.0%)	0 (0.0%)	6 (3.5%)	7 (4.1%)	0 (0.0%)	1 (0.6%)	2 (1.2%)	9 (5.2%)	143 (83.1%)

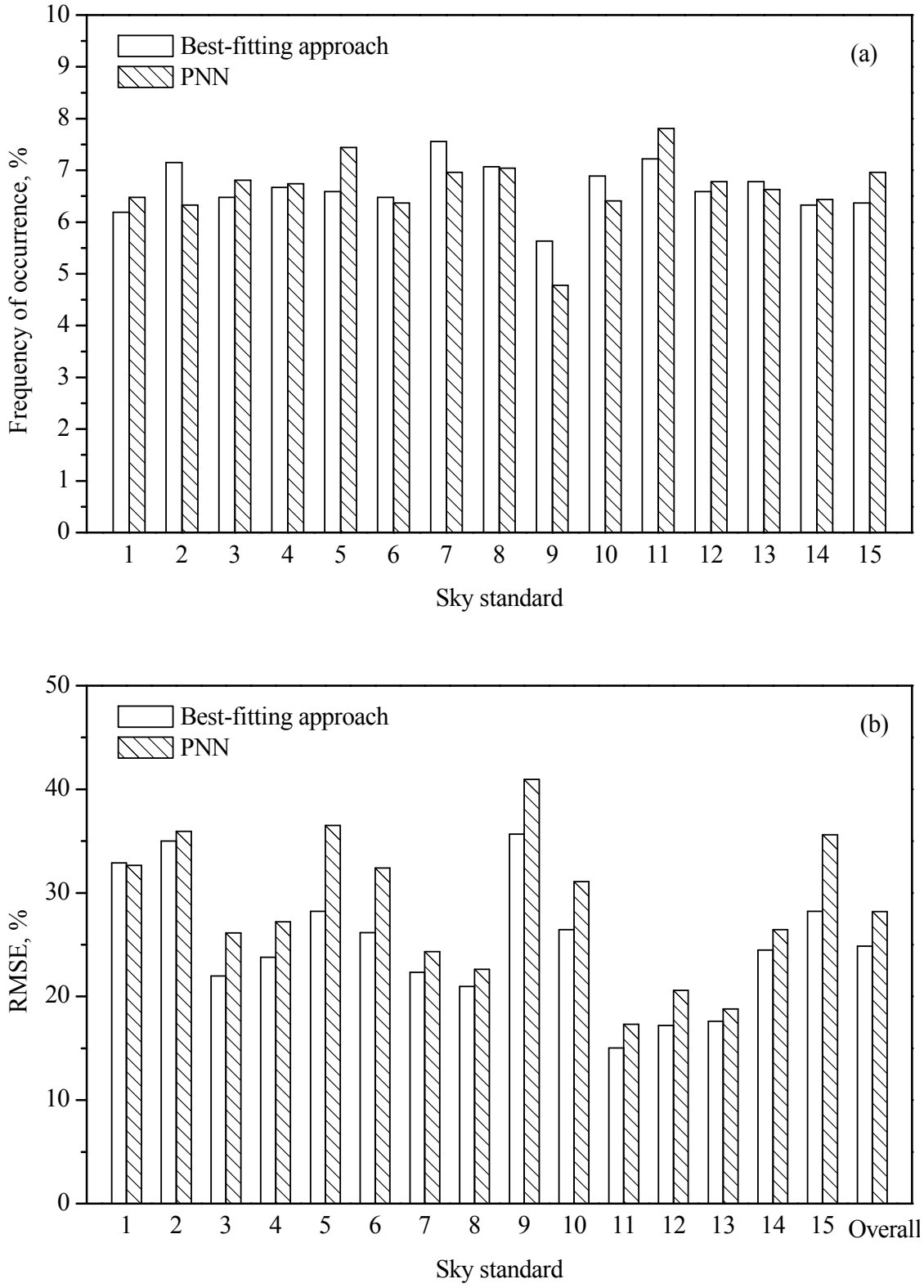

Figure 5.4 (a) Frequency of occurrence for the 15 standard skies, (b) RMSE of the luminance distribution of actual skies to the 15 standard distributions using best-fitting and PNNs approaches

5.2.3 Parametric analysis

The degree of class separability depends strongly on the selection of input parameters. The neural network regards certain parameters as important if it could give a correct response to all training samples whereas regards the other parameters as useless if the change of its values did not cause any significant improvement of the response. The performance of the PNN models for sky classification using the 5 climatic parameters had been studied in the preceding section. Nonetheless, in most of the time, all the 5 climatic parameters may not be available simultaneously. As the number of input parameters decreases, the hit percentage of PNNs would decrease accordingly. It is important to explore the accuracy of PNNs for sky classification if some of the input parameters are not available. In this section, the hit percentages of PNNs with various combinations of input parameters were presented graphically. Also, the effectiveness of each parameter affecting the sky classification was investigated.

Table 5.5 suggests 4 groups of PNNs with different combinations of input parameters. Groups I and IV represent the PNN models constructed by 1 and 4 input parameters, respectively, and a total of 5 PNN models are involved in each of these groups. Groups II and III represent the PNN models constructed by 2 and 3 input parameters respectively, and a total of 10 PNN models are involved in each of these two groups. The same manners described in section 5.1.3 such as GA and majority voting were used to train and test all the 30 PNN models, with the exception that the number of input parameters was reduced instead of 5. To investigate the influences of each input variables on the performance of PNN models, the same test set and the 30 pairs of training and validation sets were provided to each PNN model with majority voting scheme for analysis.

Table 5.5 Different combinations of input parameters

PNN	Input parameters	PNN	Input parameters	PNN	Input parameters
IV.1	$\alpha_s, \frac{L_z}{D_v}, \frac{G_v}{E_v}, \frac{D_v}{E_v}$	III.6	$\alpha_s, \frac{D_v}{E_v}, T_v$	II.6	$\frac{L_z}{D_v}, \frac{D_v}{E_v}$
IV.2	$\alpha_s, \frac{L_z}{D_v}, \frac{G_v}{E_v}, T_v$	III.7	$\frac{L_z}{D_v}, \frac{G_v}{E_v}, \frac{D_v}{E_v}$	II.7	$\frac{L_z}{D_v}, T_v$
IV.3	$\alpha_s, \frac{L_z}{D_v}, \frac{D_v}{E_v}, T_v$	III.8	$\frac{L_z}{D_v}, \frac{G_v}{E_v}, T_v$	II.8	$\frac{G_v}{E_v}, \frac{D_v}{E_v}$
IV.4	$\alpha_s, \frac{G_v}{E_v}, \frac{D_v}{E_v}, T_v$	III.9	$\frac{L_z}{D_v}, \frac{D_v}{E_v}, T_v$	II.9	$\frac{G_v}{E_v}, T_v$
IV.5	$\frac{L_z}{D_v}, \frac{G_v}{E_v}, \frac{D_v}{E_v}, T_v$	III.10	$\frac{G_v}{E_v}, \frac{D_v}{E_v}, T_v$	II.10	$\frac{D_v}{E_v}, T_v$
III.1	$\alpha_s, \frac{L_z}{D_v}, \frac{G_v}{E_v}$	II.1	$\alpha_s, \frac{L_z}{D_v}$	I.1	α_s
III.2	$\alpha_s, \frac{L_z}{D_v}, \frac{D_v}{E_v}$	II.2	$\alpha_s, \frac{G_v}{E_v}$	I.2	$\frac{L_z}{D_v}$
III.3	$\alpha_s, \frac{L_z}{D_v}, T_v$	II.3	$\alpha_s, \frac{D_v}{E_v}$	I.3	$\frac{G_v}{E_v}$
III.4	$\alpha_s, \frac{G_v}{E_v}, \frac{D_v}{E_v}$	II.4	α_s, T_v	I.4	$\frac{D_v}{E_v}$
III.5	$\alpha_s, \frac{G_v}{E_v}, T_v$	II.5	$\frac{L_z}{D_v}, \frac{G_v}{E_v}$	I.5	T_v

Figure 5.5 to Figure 5.8 show the hit percentage of the PNN models based on the test set under different categories. Under group IV (Figure 5.5), the highest hit percentage of 70.7% is found when the PNN model trained by L_z/D_v, G_v/E_v, D_v/E_v and T_v (denoted as IV.5), meaning that PNN IV.5 is the most accurate model for sky classification as the number of input parameters are limited to 4. Besides, the results also indicated that there is only slight decreases of the hit percentage (around 0.1%) as α_s is omitted. This implied that α_s is the least significant parameter and the impacts of α_s are relatively small when the skies are classified by PNNs. On the contrary, the PNN model with the combination of α_s, G_v/E_v, D_v/E_v and T_v (denoted as IV.4) has the highest misclassification rate of 43.2% and hence L_z/D_v is the critical parameter to recognize the standard sky types. Neglecting L_z/D_v as an input parameter for the PNNs would cause significant degradation of hit percentage from 70.7 to 56.8%. The hit percentage for the remaining 3 PNN models (IV.2, IV.3 and IV.4) fall within a certain range, from just over 63 to 65%, meaning that the performance of the PNN models are affected by the corresponding parameters. A detailed examination of Figure 5.5 revealed that D_v/E_v is the second critical parameter for sky classification, followed by T_v and G_v/E_v.

Further studies (Figure 5.6 and Figure 5.7) on using 2 and 3 variables as the input parameters of the PNNs support the arguments that α_s and L_z/D_v are the least and most significant parameter for sky classification, respectively. No matter what the combinations of the input variables are, the PNN models give better results when L_z/D_v is involved. Once α_s is assigned as one of the input variable of the PNNs, however, the hit percentage of the corresponding PNN models is relatively lower when compared with the other PNN models without α_s under the same group.

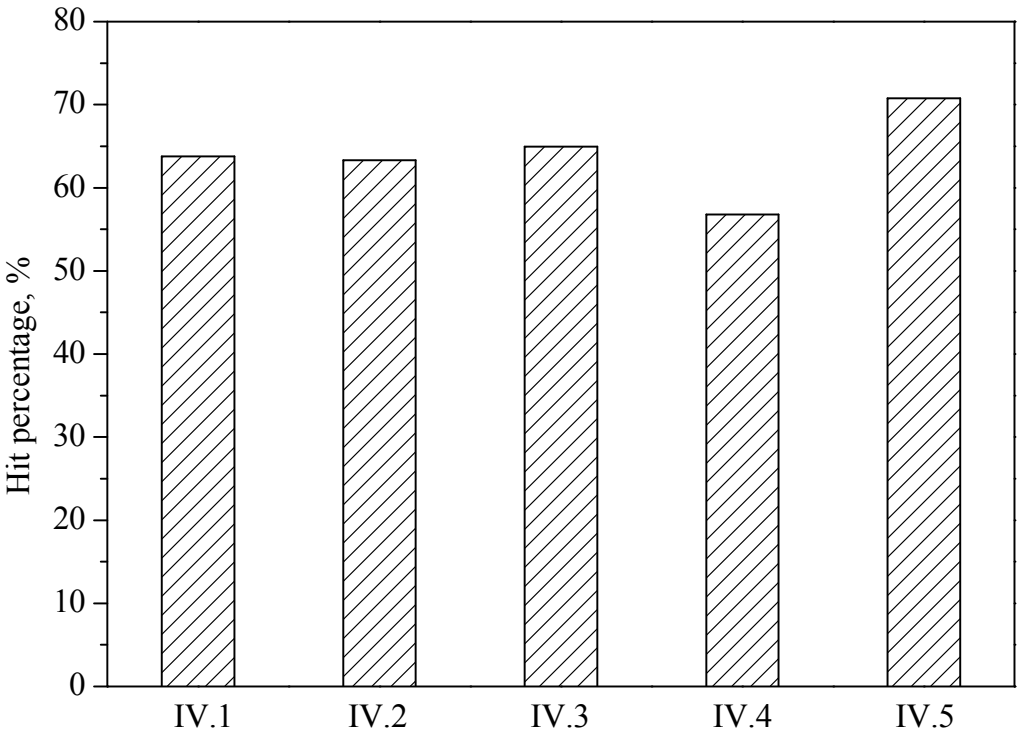

Figure 5.5 Hit percentage of the PNNs under group IV based on the test set

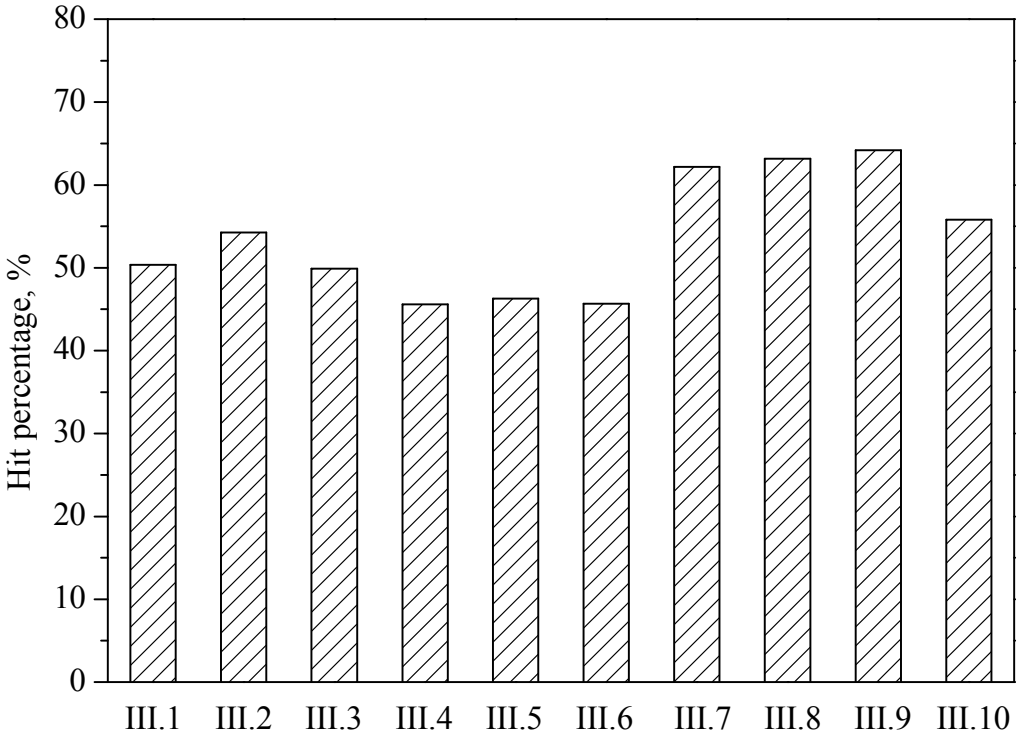

Figure 5.6 Hit percentage of the PNNs under group III based on the test set

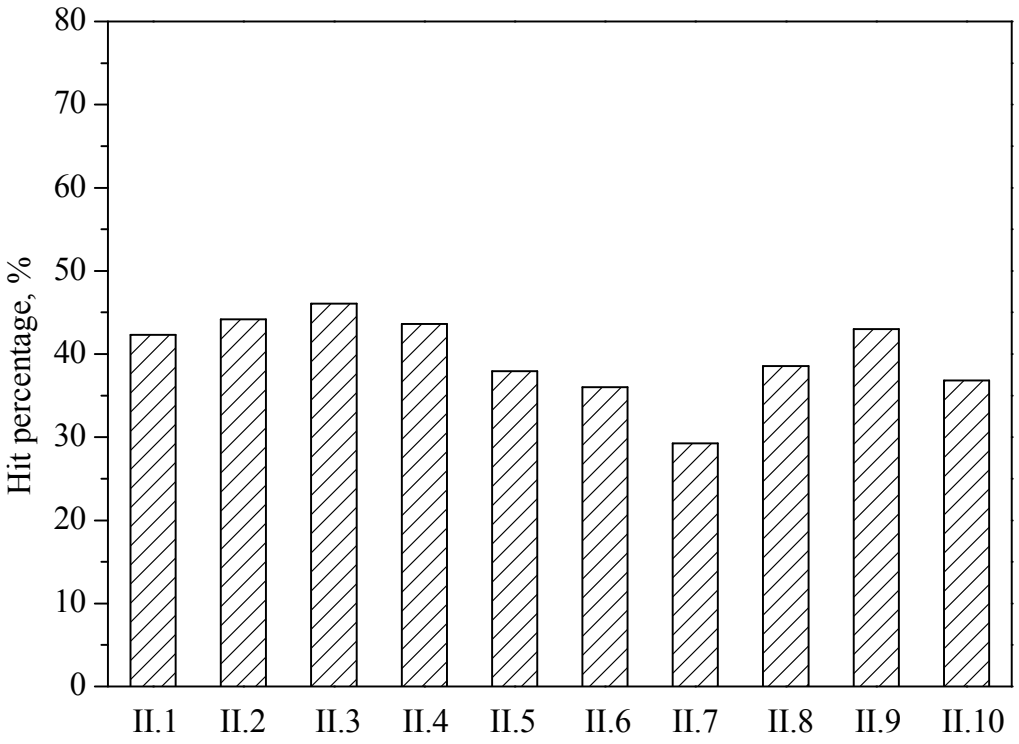

Figure 5.7 Hit percentage of the PNNs under group II based on the test set

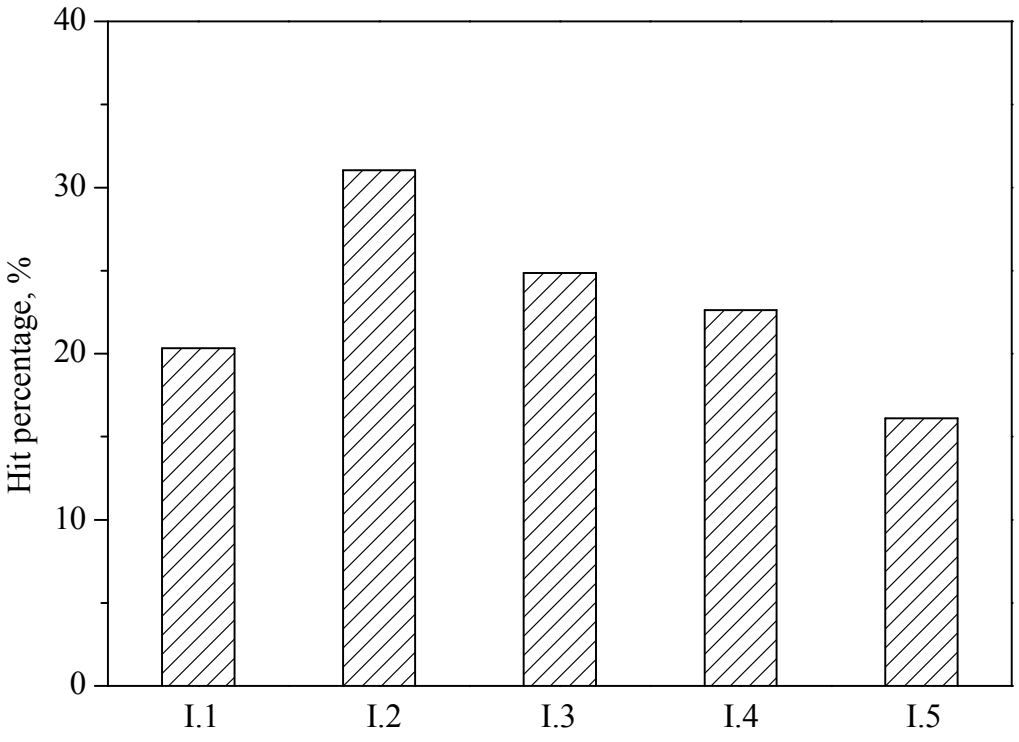

Figure 5.8 Hit percentage of the PNNs under group I based on the test set

Under group I (Figure 5.8), the best indicator is still L_z/D_v and the hit percentage of the PNN I.2 is just over 31%. However, it is interesting to note that the least accurate PNN model under group I is found when the only input parameter is T_v (denoted as I.5) instead of α_s (denoted as I.1). For PNN I.5, only 16.1% of test data are correctly classified. Obviously, only one input parameter is not sufficient to represent the characteristics of each standard sky type. The detailed results of all 30 PNN models are presented in Appendix B.

In order to study the influences of the number of input parameters on the classification results, the PNN models with the best performance among each group are compared and presented in Figure 5.9. As expected, the performance of the PNN models gives a better improvement as the number of input parameters increases. The best result is recorded when all the 5 parameters are used for the classification. As the number of input parameters is limited to 4 and 3, just a little test data (around 0.1 and 6.6% for PNN models with 4 and 3 input parameter respectively) are misclassified. However, the hit percentage of PNNs dropped rapidly to 46 and 31% as the number of input parameters were eliminated to 2 and 1 respectively.

5.3 Comparison between Climatic Parameters and PNNs Approaches

To compare the performance of the proposed climatic parameters approach with PNNs, their RMSEs of the luminance distribution of actual skies to the 15 standard skies based on 2,700 test data in section 5.1.3 are computed and presented graphically in Figure 5.10. Generally, the RMSE differences between the two approaches are small except the standard sky numbers 7, 8 and 11, meaning that the performances of the 2 approaches for classifying most of the sky types are quite similar. The peak difference of RMSE is found in standard sky number 11 with the variations ranging from 17.3% in PNNs approach to 42.6% in climatic parameters approach. It should be noted that similar RMSEs for sky numbers. 1 - 5 are observed in both approaches and thus they provide reliable ways to recognize the standard sky types under overcast sky conditions. However, particularly for the places where partly cloudy and clear skies are the prevailing sky conditions, PNN model should be used. The overall RMSEs for all 15 standard skies by PNNs approach are 28.2% which is 7% smaller than that by the climatic parameters approach.

A number of approaches suggested in this study provide simple and convenient processes to identify the 15 CIE standard skies. The selection of those approaches for practical use depends on the requirements of the users such as the level of accuracy and details. Sky classification using climatic parameters approach is a quick and simple way to be used on site when the more probable misclassifications are not considered as a serious drawback. On the contrary, ANNs is a precise and flexible tool and should be applied in detailed sky classification for daylighting design and simulation.

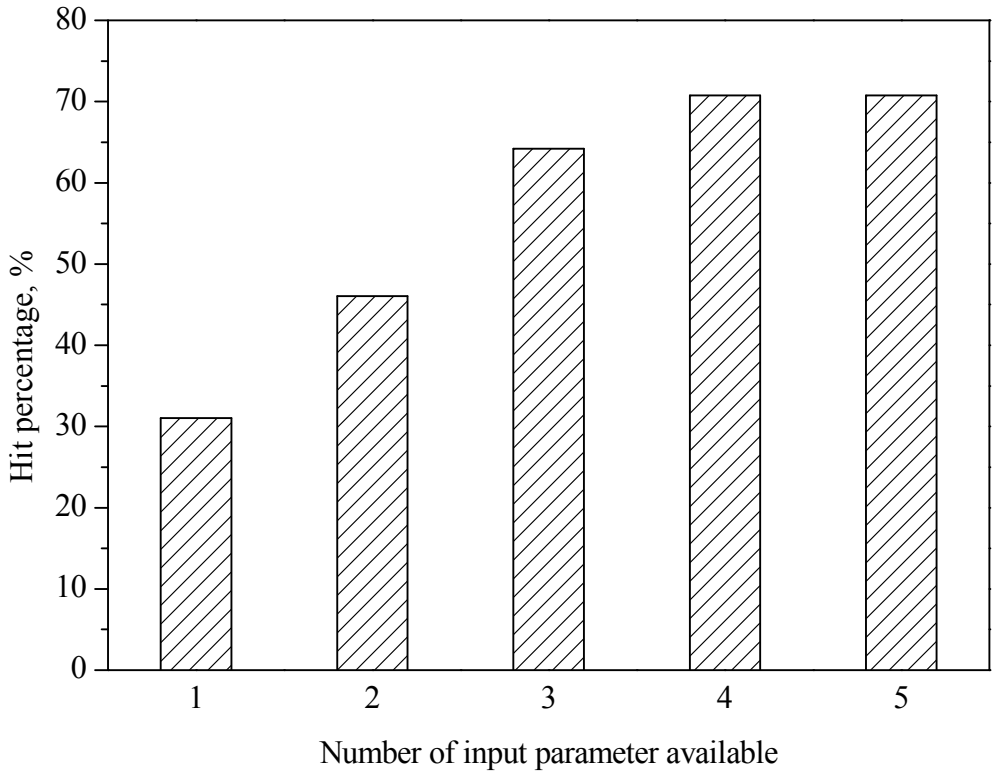

Figure 5.9 Hit percentage of the PNNs against number of sky types available

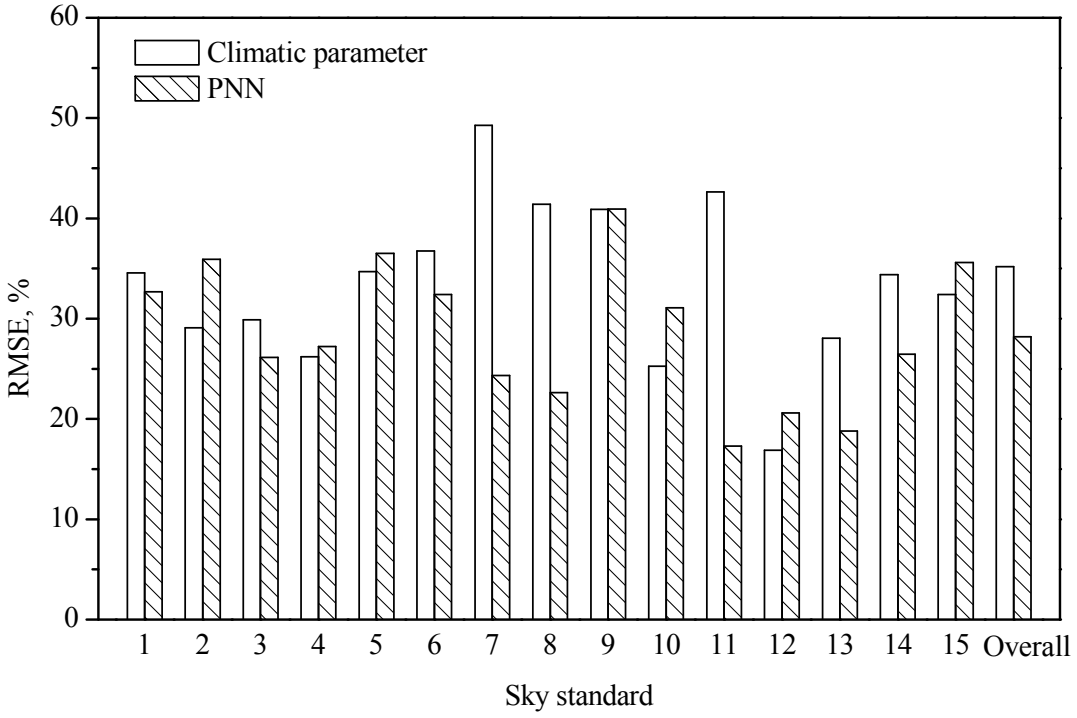

Figure 5.10 RMSE of luminance distribution of actual skies to the 15 standard skies using the climatic parameter and PNN approaches based on the test set

5.4 Summary

An evaluation of the performance of the PNNs for sky classification was carried out. It has been shown that the PNNs have the best performance for interpreting clear skies, followed by overcast and partly cloudy skies. To assess the influence of each input parameter on the classification results, parametric analysis was performed. The results concluded that L_v/D_v was the most important parameter to categorize the sky types, followed by D_v/E_v, T_v, G_v/E_v, and α_s. As the number of input parameters increases, the prediction capability of the PNN increases. At the end of this chapter, a comparative assessment between the climatic parameters and PNNs approaches was explored.

Chapter 6

CONCLUSIONS AND RECOMMENDATIONS

In this thesis, the significance of daylight and sky luminance distributions was illustrated. The daylight climate of Hong Kong, including frequency distributions and seasonal trends of outdoor illuminance, was indicated by the daylight measurement. Based on the 15 CIE sky standards, the sky distributions of Hong Kong can be identified using statistical method. Five synoptic climatic parameters were selected after considering their strengths and limitations for sky classification. Last but not least, numerous sky classification approaches were developed and evaluated. This chapter summarizes the major findings of the research, explores the limitations of the study, and provides some recommendations for future investigation on sky classification.

6.1 Summary of Major Findings

Major research findings for daylight measurement, climatic parameter selection, Hong Kong sky conditions, and the proposed sky classification approaches are summarized as follows:

6.1.1 Daylight measurement

The measured data recorded at City University of Hong Kong from January 1999 to December 2005 were collected for analysis. Due to the instrumentation malfunction, power failure and sensor calibration, considerable efforts were made to obtain a continuous record of data. After the quality control tests and diffuse illuminance corrections, approximately 68,000 10-minute readings were retained to form a reliable database.

Based on this 68,000 data samples, several analyses were conducted using both the horizontal (global and diffuse) and vertical illuminance (north, east, south and west). Frequency distributions and seasonal trends of the outdoor illuminance were computed statistically and presented graphically. The horizontal global illuminance levels at solar noon were found from 49klux in January to 66klux in July whereas the maximum and minimum horizontal diffuse illuminance around solar noon was 37klux in May and 26klux in December, respectively. For the vertical daylight illuminance, it has been found that there is a strong relationship between the variations of the vertical illuminance and the solar position. The outdoor illuminance on a vertical surface facing the sun would be the highest near sunrise and sunset hours. Moreover, the sky luminance distributions for the 3 general sky conditions (i.e. overcast, partly cloudy and clear sky) and their general features as well as characteristics were discussed.

6.1.2 Climatic parameter selection

The sky luminance distributions were caused by various factors including the solar position, turbidity and cloud amount. Climatic parameter analysis could help the identification of the sky patterns. In this study, five common climatic variables including α_s, L_z/D_v, G_v/E_v, D_v/E_v and T_v under the typical sky conditions viz. overcast, partly cloudy and clear skies were analyzed. Their characteristics, strengths and limitations in terms of sky classification were illustrated. The choice of climatic data relies on their availability, accuracy and suitability. α_s can be either measured or computed easily and it has great effects for the solar energy reaching the earth surface. However, α_s alone may not be a good factor to interpret standard skies. Although many researchers suggested that L_z/D_v is the most appropriate parameter to identify the 15 standard skies, ambiguous results may be caused when using L_z/D_v ratio to identify the sky for place where high α_s dominates.

G_v/E_v is a widely used index to interpret the sky clearness since it relies only on one measured parameter, G_v/E_v which is measured by many meteorological stations. D_v/E_v was regarded as an appropriate climatic parameter to indicate the sky condition during overcast and cloudy situations with the absence of sun. Nevertheless, the major drawback of using G_v/E_v and D_v/E_v for sky classification was that there is no clear-cut value to represent the 15 standard sky types. T_v is the best descriptor to roughly classify all clear sky types because T_v is independent of α_s and its daily course is practically stable. However, without any direct component, T_v cannot be computed accordingly and thus T_v should not be used to define the standard sky types under overcast sky conditions.

6.1.3 Hong Kong sky conditions

An analysis of the CIE standard skies against measured sky luminance data in Hong Kong was conducted. A total of 68,000 10-minute sky luminance data were carried out for the study. Based on the best-fitting approach, the appropriate sky types for subtropical Hong Kong were identified. The statistical analysis revealed that the 15 CIE standard skies are adequate to categorize the general sky conditions in Hong Kong. The overcast and clear skies contribute about 78% of the Hong Kong sky conditions. The RMSE value for the complete set of the 15 standard skies was 23.8% using best-fitting approach.

It has been found that some standard skies appear infrequently and their exclusion can simplify the analysis without significantly lowering the overall accuracy. Individual standard skies with high frequency of occurrence and low RMSE are appropriate to represent the prevailing sky conditions for a given location. Therefore, a subset of 6 prevailing sky standards (sky numbers 1, 3, 6, 7, 8 and 13) representing the actual sky conditions in Hong Kong were selected by systematical approach. The overall RMSE for the reduced set of the 6 prevailing sky standards in Hong Kong is 25% which was 1.2% greater than the complete set of standard skies.

The study was further extended to select 3 sky types (sky numbers 1, 8 and 13) for representing the overcast, partly cloudy and clear skies of Hong Kong. The overall RMSE for the subset of 3 standard skies is 27% which is 3.2 and 2% greater than the results obtained from the complete set and 6 prevailing set of standard skies, respectively.

6.1.4 Proposed sky classification approaches

The classification of the CIE standard skies using sky luminance and other climatic parameters measured in Hong Kong was presented. The sky identification should not only determine the appropriate sky type but also the frequency of occurrence in that particular location. In total, three approaches were evaluated and to discriminate the 15 standard skies: (i) Bartzokas et al. method, (ii) Climatic parameter, and (iii) ANNs. The performance of the each approach was evaluated against the measured data in terms of RMSE. A comparative assessment of the best-fitting and the proposed methods against sky luminance data scanned between 1999 and 2005 was reported.

Bartzokas et al. method

To categorize the sky type for each set of data, the L_z/D_v value at a particular α_s was compared with the L_z/D_v values of the 15 theoretical curves. Around 50% of the data were rejected due to the clear sky conditions at high α_s. Similar to the results for Bratislava in winter [Bartzokas et al., 2003], more normal cases were observed under overcast and clear skies. The RMSE varies from 24.5% for sky standard 13 to 51.3% for sky standard 9. The overall RMSE was 34.6% which is 10.8% greater than the results computed by the best-fitting method.

Climatic parameter approach

A study on identification of the 15 standard skies and 6 prevailing skies was carried out using various climatic parameters. All the measured data were initially interpreted as overcast, partly cloudy and clear sky types by the hybrid daylight-variable, L_z/D_v-G_v/E_v. Specific classification of the 15 standard skies and 6 prevailing skies was performed

based on the characteristics and features of each sky type. Different ranges of climatic parameters were proposed and the choice was relied on their strengths and limitations to single out each of the sky types.

Without further data rejection, the frequency of occurrence pattern was generally similar to that using the best-fitting approach and the overall RMSE for sky identification of 15 standard skies and the 6 prevailing skies was found to be 32.4% (8.6% greater than the best-fitting method) and 30.0% (6.2% greater than best-fitting method), respectively.

ANNs approach

The potential of using PNNs with majority voting technique in developing a model for sky classification was demonstrated. By using the five climatic variables namely α_s, L_v/D_v, G_v/E_v, D_v/E_v and T_v as the input parameters of the PNN, the 3 general sky conditions and the 15 CIE standard skies were classified. Using the PNN with majority voting techniques, about 89.2% of data sample was correctly categorizing for the 3 typical sky conditions. For identification of the 15 CIE standard skies, the hit percentage reduced, ranging from 49.1% for sky standard 6 to 86.7% for sky standard 11. In general, the PNN models show the best performance for interpreting clear skies, then overcast skies, and finally partly cloudy skies. The frequency of occurrence and the corresponding RMSE for the 2,700 test data were also determined with the best-fitting and PNN approaches. The frequency distributions were quite even and similar patterns were found between the 2 methods. The RMSE for the 15 standard skies based on PNN technique was 28.2% which was 3.3% more than that via the best-fitting approach.

Parametric analysis was also carried out to explore the influence of each parameter on sky classification. The results concluded that L_v/D_v was the most important parameter to categorize the sky types, followed by D_v/E_v, T_v, G_v/E_v, and α_s. Eliminating L_v/D_v as an input parameter of the PNN caused a significant decrease on classification accuracy, from 70.7 to 56.8%. Moreover, the performance of the PNN was affected by varying the number of input parameters. The more input parameters were considered, the higher the prediction capability of the PNN.

6.2 Limitations of the Study

In this study, two approaches based on different techniques are purposed for sky classification. However, there are several limitations that may restrict the use of those models. They are listed as follows:

i. The sky luminance distributions vary from time to time. However, the sky scanner spent about 4 minutes to complete one scanning period and hence the luminance of the 145 sky patches were not measured at the same time. The measurement errors were induced once substantial variations in sky luminance was occurred within each scan.

ii. The proposed set of 6 prevailing skies was based on the choice of the threshold values of the ratios of the frequency of occurrence to the RMSE for each sky type. These threshold values may be site dependent. Therefore, the proposed sky classification strategy for the set of 6 prevailing skies may not be suitable for places where its daylight climate was differed with Hong Kong.

iii. The performance of PNN depends on the sizes of the training, validation and test sets. However, there is no formal scientific theory to determine the required sizes of the three data sets. For different applications of the PNN, researchers recommended different criteria to partition of the database. Therefore, the development and use of the neural networks may require more expertise and experiences than the traditional classification techniques.

6.3 Suggestions for Future Study

To give a comprehensive classification of standard skies, further studies and investigations on various directions are recommended and pointed out as follows:

i. Further work should be concentrated on the selection of climatic parameters such that accuracy of the proposed approaches can be enhanced. Primary parameters include sky clearness index and diffuse fraction index. Secondary auxiliary include sunshine hour and cloud cover.

ii. Another potential extension of the present research is the study on the sky distributions over the seasonal and diurnal change. Daylighting design and indoor illuminance level depend on the sky distributions. The investigation on average fluctuations of the sky patterns over a day and season are useful for improving the performance of the simulation program as well as designing the energy efficient scheme.

iii. Another area that researchers can pursue is the integration of feature extraction techniques and PNN. Data reduction or feature extraction technique such as principle component analysis (PCA) can reduce the dimensionality of a data sample while retaining most of the information of the input parameters. It is desirable to minimize the number of input parameters of the PNN because the major drawbacks of PNN are the large demand of computational memory and low processing speed when the size of the training set is large.

REFERENCES

Aoki Y, Taniguchi T and Irikura T (1997). Analysis of space distribution of scattered daylight by the Monte Carlo Method giving consideration to atmospheric particles. Journal of Light and Visual Environment 21(2):14-22.

Aries MBC and Newsham GR (2008). Effect of daylighting saving time on lighting energy use: A literature review. Energy Policy 36(6):1858-1866.

Barbaro S, Cannata G, Coppolino S, Leone C and Singagra E (1981). Correlation between relative sunshine and state of the sky. Solar Energy 26(6):537-550.

Bartzokas A, Darula S, Kambezidis HD and Kittler R (2003). Sky luminance distribution in Central Europe and the Eastern Mediterranean area during the winter period. Journal of Atmospheric and Solar-Terrestrial Physics 65(1):113-119.

Bartzokas A, Kambezidis HD, Darula S and Kittler R (2005). Comparison between winter and summer sky-luminance distribution in Central Europe and in the Eastern Mediterranean. Journal of Atmospheric and Solar-Terrestrial Physics 67(7):709-718.

Basheer IA and Hajmeer MN (2000). Artificial neural networks: fundamentals, computation, design and application. Journal of Microbiological Method 43(1):3-31.

Bellia L, Cesarano A, Minichiello F and Sibilio S (2000). De-Light: a software tool for the evaluation of direct daylighting illuminances both indoors and outdoors – comparison with Superlite 2.0 and Lumen Micro 7.1. Building and Environment 35(4):281-295.

Bodart M and De Herde A (2002). Global energy savings in offices buildings by the use of daylighting. Energy and Buildings 34(5):421-429.

Cacoullos T (1966). Estimation of multivariate density. Annals of the Institute of Statistical Mathematics 18(2):179-189.

Calderon TG and Cheh JJ (2002). A roadmap for future neural networks research in auditing and risk assessment. International Journal of Accounting Information Systems 3(4):203-236.

Cam E, Arcaklioglu E, Cavusoglu A and Akbiyik B (2005). A classification mechanism for determining average wind speed and power in several regions of Turkey using artificial neural networks. Renewable Energy 30(2):227-239.

Chen WC and Hsu SW (2007). A neural network approach for an automatic LED inspection system. Expert System with Applications 33(2):531-537.

Chirarattananon S and Chaiwiwatworakul P (2007). Distributions of sky luminance and radiance of North Bangkok under standard distributions. Renewable Energy 32(8):1328-1345.

CIE (1973). CIE publication no.22 - Standardization of luminance distribution on clear skies. Paris: Commission International Commission of Illumination.

CIE (1994). CIE technical report (CIE 108-1994) – Guide to recommended practice of daylight measurement. Vienna: International Commission of Illumination.

CIE (2003). Spatial distribution of daylight – CIE standard general sky. Vienna: International Commission of Illumination Central Bureau.

Darula S and Kittler R (2004a). New trends in daylight theory based on the new ISO/CIE sky standard: 1.Zenith luminance on overcast skies. Building Research Journal 52(3):181-197.

Darula S and Kittler R (2004b). New trends in daylight theory based on the new ISO/CIE sky standard: 2.Typology of cloudy skies and their zenith luminance. Building Research Journal 52(4):245-255.

Darula S and Kittler R (2005). New trends in daylight theory based on the new ISO/CIE sky standard: 3.Zenith luminance formula verified by measurement data under cloudless skies. Building Research Journal 53(1):9-31.

Darula S, Kittler R and Kmeto P (2001). New CIE general sky defining luminance distributions. Proceedings of the Sustainable Building & Solar Energy Conference, Brno. p.52-54.

Doulos L, Tsangrassoulis A and Topalis F (2008). Quantifying energy savings in daylight responsive systems: the role of dimming electronic ballasts. Energy and Buildings 40(1):36-50.

Drummond AJ (1956). On the measurement of sky radiation. Archiv fur Meteorologie, Geophysik und Bioklimatologie 7(B):413-436.

EKO (1998). Instruction manual – sky scanner MS-300LR. Japan: EKO.

Enarun D and Littlefair P (1995). Luminance models for overcast skies: assessment using measured data. Lighting Research and Technology 27(1):53-58.

Gaganis C, Pasiouras F and Doumpos M (2007). Probabilistic neural networks for the identification of qualified audit opinions. Expert Systems with Applications 32(1):114-124.

Galasiu AD and Veitch JA (2006). Occupant preferences and satisfaction with the luminous environment and control systems in daylit offices: a literature review. Energy and Buildings 38(7):728-742.

Guler I and Ubeyli ED (2007). Implementing wavelet/probabilistic neural networks for Doppler ultrasound blood flow signals. Expert Systems with Applications 33(1):162-170.

Hajmeer M and Basheer I (2002). A probabilistic neural network approach for modeling and classification of bacterial growth/no-growth data. Journal of Microbiological Methods 51(2):217-226.

Hansen J, McDonald J and Slice J (1992). Artificial intelligence and generalized qualitative response models: an empirical test on two audit decision making domains. Decision Science 23(3):708-723.

Hay JE (1993). Solar radiation data: validation and quality control. Renewably Energy 3(4-5):349-355.

Hayashi Y, Setiono R and Yoshida K (2000). A comparison between two neural network rule extraction techniques for the diagnosis of hepatobiliary disorders. Artificial Intelligence in Medicine 20(3):205-216.

Haykin S (1999). Neural networks, a comprehensive foundation. New Jersey: Prentice-Hall.

Heschong L (2002). Daylighting and human performance. ASHRAE Journal 44(8):65-67.

Igawa N, Nakamura H, Matsuzawa T, Koga Y, Gota K and Kojo S (1997). Sky luminance distribution between two CIE skies. Proceedings of Lux Pacifica, Nagoya. p.E7-E18.

Ihm P, Nemri A and Krarti M (2008). Estimation of lighting energy savings from daylighting. Building and Environment. In press.

Jenkins D and Newborough M (2007). An approach for estimating the carbon emissions associated with office lighting with a lighting contribution. Applied Energy 84(6):608-622.

Jeter SM and Balaras CA (1990). Development of improved solar radiation models for predicting beam transmittance. Solar Energy 44(3):149-156.

Kalatzis I, Piliouras N, Glotsos D, Ventouras E, Papageorgiou C, Rabavilas A, Soldatos C and Cavouras D (2005). Identifying differences in the P600 component of ERP-signals between OCD patients and controls employing a PNN-based majority vote classification scheme. Proceedings of the 2005 IEEE Engineering in Medicine and Biology Conference, Shanghai. p.3994-3997.

Kambezidis HD, Katevatis EM, Petrakis M, Lykowdis S and Asimakopoulos DN (1998). Estimation of the Link and Unsworth-Monteith turbidity factors in the visible spectrum: Application for Athens, Greece. Solar Energy 62(1):39-50.

Karayel M, Navvab M, Ne'eman E and Selkowitz S (1984). Zenith luminance and sky luminance distributions for daylighting calculations. Energy and Buildings 6(3):283-291.

Kasper S, Wehr TA, Bartko JJ, Gaist PA. Rosenthal NE (1989). Epidemiological findings of seasonal changes in mood and behavior. Archives of General Psychiatry 46(9):823-833.

Kasten F and Young AT (1989). Revised optical air mass tables and approximation formula. Applied Optics 28(22):4735-4738.

Kim D, Kim DH and Chang S (2008). Application of probabilistic neural network to design breakwater armor blocks. Ocean Engineering 35(3-4):294-300.

Kipp & Zonen. User manual – shadow ring CM121. Delft: Kipp & Zonen.

Kittler R (1967). Standardization of outdoor conditions for the calculation of daylight factor with clear skies. Proceeding of CIE Intercessional Conference on Sunlight in Buildings, Newcastle. P.273-285.

Kittler R and Darula S (1998). Parameterization problems of the very bright cloudy sky conditions. Solar Energy 62(2):93-100.

Kittler R and Darula S (2000). Determination of sky types from global illuminance. Lighting Research and Technology 32(4):187-193.

Kittler R and Darula S (2002). Parametric definition of the daylight climate. Renewable Energy 26(2):177-187.

Kittler R, Darula S and Perez R (1998). A set of standard skies. Bratislava: Polygrafia.

Kittler R, Darula S and Perez R (1999). Advantages of new sky standards: more realistic modeling of daylight conditions in energy and environmental studies. International Journal of Energy, Environment and Economics 8(1):65-71.

Kittler R, Hayman S, Ruck N and Julian W (1992). Daylight measurement data: Methods of evaluation and representation. Lighting Research and Technology 24(4):173-187.

Kittler R, Perez R and Darula S (1997a). A new generation of sky standards. Proceedings of the LUX Europa Conference, Amsterdam. p.359-373.

Kittler R, Perez R and Darula S (1997b). Sky classification respecting energy-efficient lighting, glare and control needs. Journal of Illuminating Engineering Society 26(1):57-68.

Kobav MB and Bizjak G (2003). Simulating and calculating sky luminance distribution with CIE standard DS 011.2. Proceedings of the IEEE Region 8 EUROCON, Slovenia. p.516-520.

Kudish AI and Ianetz A (1993). Analysis of diffuse radiation data for beer sheva: measured (shadow ring) versus calculated (global-horizontal beam) values. Solar Energy 51(6):495-503.

Lam WMC (1985). Sunlighting as form-givers for architecture. New York: McGraw-Hill.

Lam JC, Li DHW and Cheung SO (2003). An analysis of electricity end-use in air-conditioned office buildings in Hong Kong. Building and Environment 38(3):493-498.

Lau SY (1989). Technical note no.81 – global solar radiation in Hong Kong. Hong Kong: Hong Kong: Royal Observatory.

LeBaron BA, Michalsky JJ and Perez R (1990). A simple procedure for correcting shadowband data for all sky conditions. Solar Energy 44(5):249-256.

LeBaron BA, Peterson WA and Dirmhirn I (1980). Corrections for diffuse irradiance measured with shadowbands. Solar Energy 25(1):1-13.

Lee EWM, Lee YY, Lim CP and Tang CY (2006). Application of a noisy data classification technique to determine the occurrence of flashover in compartment fires. Advanced Engineering Informatics 20(2):213-222.

Leslie RP (2003). Capturing the daylight divided in buildings: why and how?. Building and Environment 38(2):381-385.

Li DHW (2007). Daylight and energy implications for CIE standard skies. Energy Conversion and Management 48(3) 745-755.

Li DHW, Cheung GHW and Cheung KL (2006a). Evaluation of simplified procedure for indoor daylight illuminance determination against data in scale model measurements. Indoor Built and Environment 15(3):213-223.

Li DHW, Lam TNT and Wong SL (2006b). Lighting and energy performance for an office using high frequency dimming controls. Energy Conversion and Management 47(9-10):1133-1145.

Li DHW and Lam JC (2001). An analysis of climatic parameters and sky condition classification. Building and Environment 36(4):435-445.

Li DHW and Lau CCS (2007). An analysis of nonovercast sky luminance models against Hong Kong. Journal of Solar Energy Engineering 129(4):486-493.

Li DHW, Lau CCS and Lam JC (2001). Evaluation of overcast sky luminance models against measured Hong Kong Data. Applied Energy 70(4):321-331.

Li DHW, Lau CCS and Lam JC (2003). A study of 15 sky luminance patterns against Hong Kong data. Architectural Science Review 46(1):61-68.

Li DHW, Lau CCS and Lam JC (2004a). Overcast sky conditions and luminance distribution in Hong Kong. Building and Environment 39(1):101-108.

Li DHW, Lau CCS and Lam JC (2004b). Standard skies classification using common climatic parameters. Journal of Solar Energy Engineering 126(3):957-964.

Li DHW, Lau CCS and Lam JC (2005). Predicting daylight illuminance on inclined surfaces using sky luminance data. Energy – The International Journal 30(9):1649-1665.

Li DHW and Tang HL (2008). Standard skies classification in Hong Kong. Journal of Atmospheric and Solar-Terrestrial Physics 70(8-9):1222-1230.

Lim CP and Harrison RF (1997). An Incremental adaptive network for on-line supervised learning and probability estimation. Neural Networks 10(5):925-939.

Lim CP, Quek SS and Peh KK (2003). Predicting of drug release profiles using an intelligent learning system: an experimental study in transdermal iontophoresis. Journal of Pharmaceutical and Biomedical Analysis 31(1):159-168.

Lin W and Lu E (1992). Annual and seasonal global solar radiation climates in Yunnan, China. Energy Conversion and Management 33(11):1097-1099.

Littlefair PJ (1989). Correcting for the shade ring used in diffuse daylight and radiation measurements. Proceedings of the Daylight and Solar Radiation Measurements CIE – WMO Symposium, Berlin. p.233-241.

Littlefair PJ (1994). The luminance distributions of clear and quasi-clear skies. Proceedings of the CIBSE National Lighting Conference, Cambridge, p.267-283.

Long CN and Ackerman TP (2000). Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud-effects. Journal of Geophysical Research 105(15):609-615.

Longmore J (1975). Daylighting: a current view. Light and Lighting 68(3):113-119.

Lundt L (2004). Modafinil treatment in patients with seasonal affective disorder winter depression: an open-label pilot study. Journal of Affective Disorders 81(2):173-178.

Magnusson A (2000). An overview of epidemiological studies on seasonal affective disorder. Acta Psychiatrica Scandinavica 101(3):176–184.

Mao KZ, Tan KC and Ser W (2000). Probabilistic neural network structure determination for pattern classification. IEEE Transactions on neural networks 11(4):1009-1016.

Markou MT, Kambezidis HD, Bartzokas A, Katsoulis BD and Muneer T (2005). Sky type classification in Central England during winter. Energy 30(9):1667-1674.

Markou MT, Bartzokas A and Kambezidis HD (2007). A new statistical methodology for classification of sky luminance distributions based on scan data. Atmospheric Research 86(3-4):261-277.

Masters T (1994). Practical Neural Network Recipes in C++. Boston: Academic Press.

Mood AM and Graybill FA (1962). Introduction to the theory of statistics. New York: McGraw-Hill.

Moon P and Spencer DE (1942). Illumination from a non-uniform sky. Illuminating Engineering 37(10):707-726.

Nabhani F and Shaw T (2002). Performance analysis and optimization of shape recognition and classification using ANN. Robotics and Computer-Integrated Manufacturing 18(3-4):177-185.

Nakamura H, Oki M and Hayashi Y (1985). Luminance distribution of intermediate sky. Journal of Light and Visual Environment 9(1):6-13.

Nakamura H, Oki M and Iwata T (1987). Mathematical description of the intermediate sky. Proceeding of 21st CIE session, Venice. p.230-231.

Nasseri M, Asghari K and Abedini MJ (2008). Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. Expert Systems with Applications 35(3):1415-1421.

Navvab M, Karayel M, Ne'eman E and Selkovitz S (1984). Analysis of atmospheric turbidity for daylight calculations. Energy and Buildings 6(3):293-303.

NeuroShell 2 (2000). NeuroShell 2 – Release 4.0. Maryland: Ward System Group.

Ng E, Cheng V, Gadi A, Mu J, Lee M and Gadi A (2007). Defining standard skies for Hong Kong. Building and Environment 42(2):866-876.

Onaygil S and Guler O (2003). Determination of the energy saving bty daylight responsive lighting control system with an example from Istanbul. Building and Environment 38(7):973-977.

Osowski S, Markiewicz T and Tran Hoai L (2008). Recognition and classification system of arrhythmia using ensemble of neural networks. Measurement 41(6):610-617.

Painter HE (1981). The shade ring correction for diffuse irradiance measurements. Solar Energy 26(4):361-363.

Parzen E (1962). On estimation of a probability density function and mode. Annals of Mathematical Statistics 33(3):1065-1076.

Peacock JE (1978). Technical Note No.14 – Solar data for Hong Kong. Hong Kong: Royal Observatory Hong Kong.

Perez R, Ineichen P and Seals R (1990). Modeling daylight availability and irradiance components from direct and global irradiance. Solar Energy 44(5):271-289.

Perez R, Seals R and Michalsky J (1993). Modeling skylight angular luminance distribution from routine irradiance measurements. Journal of the Illuminating Engineering Society 22(1):10-17.

Plympton P, Conway S and Epstein K (2000). Daylighting in schools: improving student performance and health at a price schools can afford. Proceedings of the ASES Solar 2000 Passive Conference, Wisconsin. p.487-492.

Quteishat A and Lim CP (2008). A modified fuzzy min-max neural network with rule extraction and its application to fault detection and classification. Applied Soft Computing 8(2):985-995.

Ravi V, Kurniawan H, Thai PNK and Ravi Kumar P (2008). Soft computing system for bank performance prediction. Applied Soft Computing 8(1):305-315.

Reinhart C and Selkowitz S (2006). Daylighting – light, form and people. Energy and Buildings 38(7):715-717.

Rutten AJF (1994). Sky luminance research imperative for adequate control of temporary supplementary artificial lighting installations. Building and Environment 29(1):105-111.

Salchenberger LM, Cinar EM and Lash NA (1992). Neural networks: a new tool for predicting thrift failures. Decision Sciences 23(4):899-916.

Shahriar ANM and Mohit MA (2006). Frequency distribution of CIE standard general skies for Subang, Malaysia. Architectural Science Review 49(4):363-366.

Sham P (1964). Technical Note No.22 – Total solar and sky radiation in Hong Kong. Hong Kong: Royal Observatory Hong Kong.

Smiley F (1996). Students delight in daylight. International Association for Energy Efficient Lighting Newsletter 5(2):11-12.

Soler A and Gopinathan KK (2000). A study of zenith luminance on Madrid cloudless skies. Solar Energy 69(5):403-411.

Specht D (1990a). Probabilistic neural networks. Neural Networks 3(1):109-118.

Specht D (1990b). Probabilistic neural networks and the polynomial adaline as complementary techniques for classification. IEEE Transactions on Neural Networks 1(1):111-121.

Specht DF and Shapiro PD (1990). Training speed comparison of probabilistic neural networks with back-propagation networks. Proceedings of the International Neural Network Conference, Paris. p.440-443.

Specht DF and Shapiro PD (1991). Generalization accuracy of probabilistic neural networks compared with back-propagation networks. Proceedings of the International Joint Conference on Neural Networks, Washington. p.887-892.

Swingler K (1996). Applying neural networks: a practical guide. New York: Academic Press.

Tregenza PR (1987). Subdivision of the sky hemisphere for luminance measurements. Lighting Research and Technology 19(1):13-14.

Tregenza PR (1999). Standard skies for maritime climates. Lighting Research and Technology 31(3):97-106.

Tregenza PR (2004). Analysing sky luminance scans to obtain frequency distributions of CIE Standard General Skies. Lighting Research and Technology 36(4):271-281.

Tregenza PR and Sharples S (1995). New daylight algorithm. Sheffield: University of Sheffield.

Van der Meer JW (1988). Rock slopes and gravel beaches under wave attack, PhD Thesis, Delft University of Technology, Delft, Netherlands.

Vartiainen E (2000). Daylight modeling and optimization of solar facades, PhD Thesis, Helsinki University of Technology, Espoo, Finland.

Webb AR (2006). Considerations for daylighting in the built environment: Non-visual effects on light. Energy and Buildings 38(7):721-727.

Wittkopf SK, Yuniarti E and Soon LK (2006). Prediction of energy savings with anidolic integrated ceiling across different daylight climates. Energy and Buildings 38(9):1120-1129.

Younes S, Claywell R and Muneer T (2005). Quality control of solar radiation data: present status and proposed new approaches. Energy 30(9):1533-1549.

Younes S and Muneer T (2007). Clear-sky classification procedures and models using a world-wide data-base. Applied Energy 84(6):623-645.

Zaknich A (2003). Neural networks for intelligent signal processing. London: World Scientific.

Zmeureanu R and Peragine C (1999). Evaluation of interactions between lighting and HVAC systems in a large commercial building. Energy Conversion and Management 40(11):1229-1236

APPENDIX A

SUN PATH DIAGRAM FOR HONG KONG

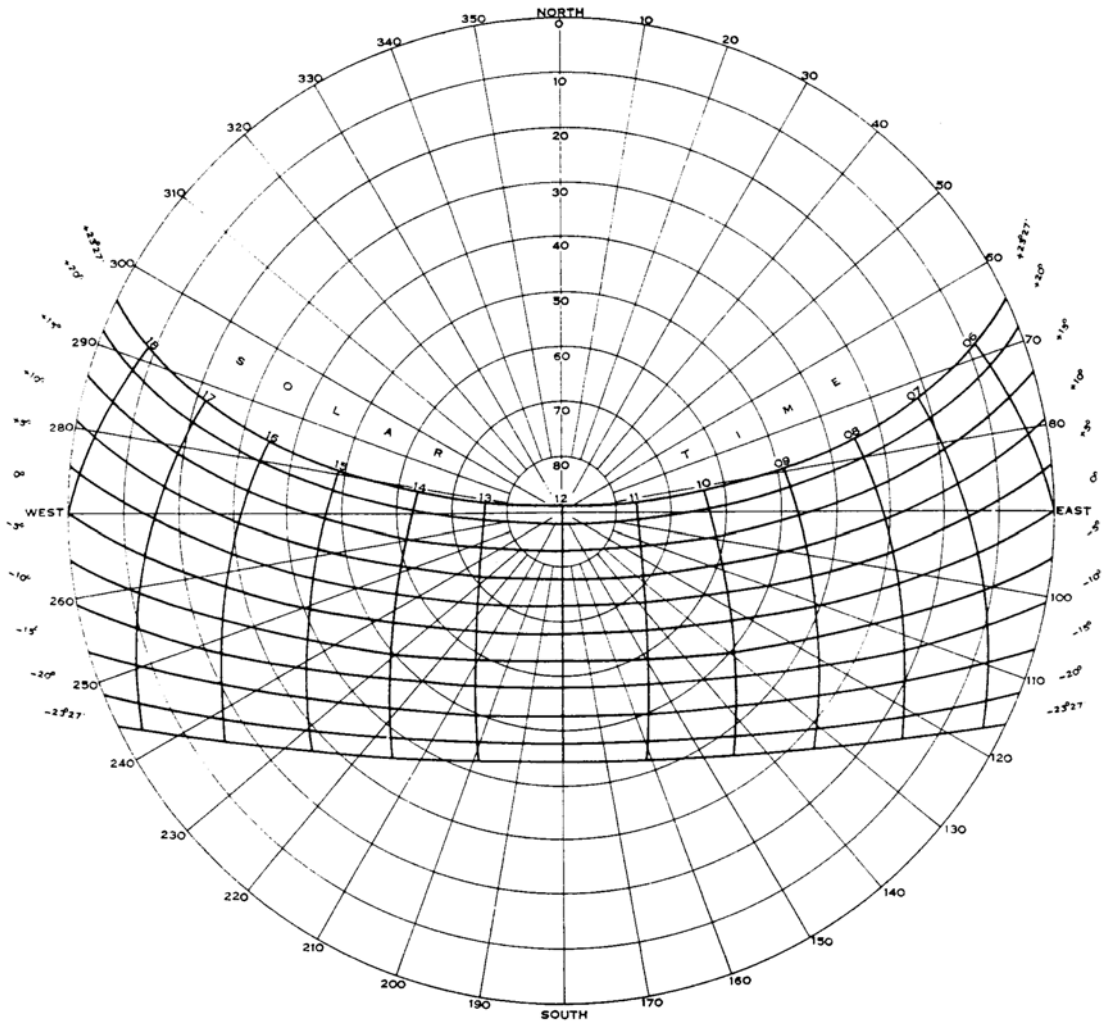

Figure A1 Sun path diagram for Hong Kong [Peacock, 1978]

APPENDIX B

RESULTS OF THE PNN CLASSIFICATION

Table B1 Overall results of the test set for PNN IV.1

PNN IV.1				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	182	6.7	33.1	65.3
2	172	6.4	34.3	58.5
3	203	7.5	27.2	60.6
4	191	7.1	28.6	53.3
5	165	6.1	37.1	50.0
6	164	6.1	33.2	41.7
7	191	7.1	26.1	55.4
8	198	7.3	22.6	71.2
9	141	5.2	42.8	46.7
10	161	6.0	28.2	47.3
11	209	7.7	18.2	82.6
12	197	7.3	22.4	81.5
13	182	6.7	20.0	81.4
14	183	6.8	27.8	83.6
15	161	6.0	30.4	75.6
Overall	2700	100.0	28.3	63.8

Table B2 Overall results of the test set for PNN IV.2

PNN IV.2				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	179	6.6	32.8	77.2
2	173	6.4	36.1	61.7
3	187	6.9	26.7	74.3
4	173	6.4	26.8	58.3
5	198	7.3	36.6	56.7
6	159	5.9	30.5	41.7
7	196	7.3	25.4	58.3
8	199	7.4	23.8	68.6
9	126	4.7	43.0	44.7
10	154	5.7	32.9	39.8
11	225	8.3	18.1	78.5
12	199	7.4	21.6	74.2
13	185	6.9	21.5	66.1
14	165	6.1	27.1	74.3
15	182	6.7	35.4	73.8
Overall	2700	100.0	28.6	63.3

Table B3 Overall results of the test set for PNN IV.3

PNN IV.3				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	183	6.8	33.9	76.6
2	159	5.9	33.8	61.1
3	195	7.2	26.7	73.7
4	195	7.2	28.8	58.3
5	187	6.9	37.2	57.3
6	162	6.0	29.2	46.3
7	177	6.6	24.7	54.4
8	211	7.8	24.0	74.9
9	122	4.5	44.3	43.4
10	150	5.6	32.6	42.5
11	214	7.9	17.8	80.0
12	194	7.2	21.0	73.0
13	190	7.0	20.0	71.0
14	185	6.9	26.7	83.6
15	176	6.5	33.6	77.3
Overall	2700	100.0	28.3	65.0

Table B4 Overall results of the test set for PNN Iv.4

PNN IV.4				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	225	8.3	35.8	80.2
2	166	6.1	36.8	42.5
3	174	6.4	32.8	57.1
4	159	5.9	34.8	38.9
5	174	6.4	37.9	41.6
6	153	5.7	35.7	35.4
7	173	6.4	28.2	43.6
8	212	7.9	28.3	55.5
9	166	6.1	44.4	53.3
10	148	5.5	38.2	37.1
11	246	9.1	23.5	74.9
12	167	6.2	28.5	66.3
13	194	7.2	22.8	74.3
14	170	6.3	25.9	74.9
15	173	6.4	30.4	80.2
Overall	2700	100.0	31.9	56.8

Table B5 Overall results of the test set for PNN IV.5

PNN IV.5				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	187	6.9	31.4	77.2
2	166	6.1	35.9	65.3
3	183	6.8	26.6	74.9
4	179	6.6	27.3	64.4
5	202	7.5	37.1	64.0
6	157	5.8	33.1	47.4
7	193	7.1	24.7	61.8
8	182	6.7	21.6	75.4
9	141	5.2	41.7	50.0
10	178	6.6	31.7	56.5
11	212	7.9	17.4	86.7
12	179	6.6	20.2	84.3
13	178	6.6	18.2	82.0
14	178	6.6	26.6	87.1
15	185	6.9	35.5	82.6
Overall	2700	100.0	28.3	70.7

Table B6 Overall results of the test set for PNN III.1

PNN III.1				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	164	6.1	31.5	49.1
2	128	4.7	35.5	40.4
3	262	9.7	31.1	54.9
4	189	7.0	26.8	43.9
5	167	6.2	39.7	38.8
6	106	3.9	27.9	23.4
7	223	8.3	27.1	54.9
8	246	9.1	24.7	71.7
9	119	4.4	45.9	45.4
10	109	4.0	32.7	23.1
11	263	9.7	18.2	76.4
12	149	5.5	26.4	34.3
13	238	8.8	19.6	62.8
14	177	6.6	28.4	66.1
15	160	5.9	29.5	67.4
Overall	2700	100.0	28.4	50.4

Table B7 Overall results of the test set for PNN III.2

PNN III.2				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	184	6.8	32.8	55.7
2	128	4.7	35.5	42.5
3	234	8.7	29.5	56.0
4	203	7.5	26.9	46.7
5	163	6.0	38.7	43.3
6	125	4.6	32.5	24.6
7	215	8.0	25.5	55.9
8	231	8.6	23.9	70.2
9	112	4.1	40.7	40.8
10	88	3.3	32.0	20.4
11	238	8.8	17.9	72.8
12	161	6.0	21.0	51.7
13	242	9.0	22.1	71.0
14	184	6.8	28.0	80.1
15	192	7.1	34.8	80.8
Overall	2700	100.0	28.3	54.3

Table B8 Overall results of the test set for PNN III.3

PNN III.3				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	166	6.1	31.3	55.7
2	162	6.0	38.4	38.9
3	153	5.7	27.6	49.7
4	234	8.7	31.4	50.6
5	158	5.9	37.4	33.1
6	137	5.1	38.5	24.6
7	234	8.7	30.7	49.0
8	251	9.3	27.3	68.6
9	121	4.5	49.5	30.9
10	101	3.7	38.4	21.0
11	268	9.9	19.7	74.4
12	174	6.4	20.7	59.0
13	190	7.0	23.1	52.5
14	198	7.3	27.1	74.3
15	153	5.7	32.5	63.4
Overall	2700	100.0	30.3	49.9

Table B 9 Overall results of the test set for PNN III.4

PNN III.4				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	250	9.3	36.4	73.7
2	125	4.6	36.9	22.8
3	218	8.1	33.6	45.1
4	192	7.1	34.6	29.4
5	165	6.1	42.5	29.8
6	127	4.7	37.5	16.6
7	178	6.6	31.6	33.3
8	228	8.4	30.8	40.3
9	124	4.6	47.7	41.4
10	95	3.5	42.3	16.1
11	305	11.3	25.3	70.3
12	149	5.5	25.9	49.4
13	215	8.0	22.8	72.7
14	162	6.0	26.7	69.6
15	167	6.2	29.1	77.9
Overall	2700	100.0	32.5	45.6

Table B10 Overall results of the test set for PNN III.5

PNN III.5				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	229	8.5	34.6	74.9
2	154	5.7	41.0	29.5
3	200	7.4	32.8	52.6
4	153	5.7	35.4	29.4
5	184	6.8	40.0	36.5
6	114	4.2	36.7	19.4
7	178	6.6	32.7	29.9
8	300	11.1	29.4	55.0
9	147	5.4	46.3	48.7
10	119	4.4	37.3	22.6
11	215	8.0	25.9	50.3
12	154	5.7	25.4	47.2
13	235	8.7	25.8	59.6
14	158	5.9	25.8	71.3
15	160	5.9	30.8	75.0
Overall	2700	100.0	32.7	46.3

Table B11 Overall results of the test set for PNN III.6

PNN III.6				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	236	8.7	34.3	73.7
2	146	5.4	39.6	29.0
3	181	6.7	32.1	48.0
4	100	3.7	36.5	19.4
5	185	6.9	40.6	32.6
6	103	3.8	38.7	14.9
7	172	6.4	32.9	29.4
8	351	13.0	31.5	56.5
9	142	5.3	46.8	47.4
10	124	4.6	39.2	25.3
11	268	9.9	28.1	60.5
12	139	5.1	28.8	46.6
13	229	8.5	25.2	61.7
14	158	5.9	27.1	69.0
15	166	6.1	30.8	76.7
Overall	2700	100.0	33.2	45.7

Table B12 Overall results of the test set for PNN III.7

PNN III.7				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	172	6.4	34.0	58.7
2	172	6.4	34.8	58.5
3	221	8.2	27.8	60.0
4	184	6.8	28.2	53.3
5	179	6.6	37.9	46.6
6	150	5.6	33.4	36.6
7	189	7.0	26.4	56.4
8	193	7.1	23.5	69.6
9	144	5.3	43.4	37.5
10	160	5.9	31.3	46.2
11	213	7.9	19.5	81.5
12	193	7.1	21.8	82.0
13	179	6.6	19.8	80.3
14	186	6.9	27.2	86.5
15	165	6.1	32.4	75.0
Overall	2700	100.0	28.9	62.2

Table B13 Overall results of the test set for PNN III.8

PNN III.8				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	192	7.1	32.5	77.8
2	167	6.2	36.7	62.2
3	184	6.8	26.7	72.0
4	171	6.3	28.5	58.9
5	211	7.8	36.6	60.1
6	157	5.8	31.0	44.6
7	191	7.1	26.3	58.8
8	193	7.1	24.4	67.0
9	121	4.5	42.6	33.6
10	155	5.7	32.2	40.9
11	223	8.3	18.2	77.9
12	191	7.1	21.7	72.5
13	186	6.9	20.9	66.1
14	167	6.2	26.6	76.6
15	191	7.1	35.9	75.6
Overall	2700	100.0	28.9	63.1

Table B14 Overall results of the test set for PNN III.9

PNN III.9				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	179	6.6	33.9	75.4
2	166	6.1	34.6	65.8
3	194	7.2	25.3	77.7
4	184	6.8	27.0	56.1
5	207	7.7	37.7	63.5
6	149	5.5	32.6	45.1
7	180	6.7	26.2	55.9
8	202	7.5	24.8	73.3
9	126	4.7	44.9	36.8
10	166	6.1	32.2	46.2
11	200	7.4	18.5	77.9
12	199	7.4	22.3	72.5
13	192	7.1	21.1	73.2
14	182	6.7	26.6	80.7
15	174	6.4	42.2	59.3
Overall	2700	100.0	29.4	64.2

Table B15 Overall results of the test set for PNN III.10

PNN III.10				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	240	8.9	36.8	77.8
2	142	5.3	40.5	40.9
3	156	5.8	30.2	53.7
4	146	5.4	35.7	38.3
5	217	8.0	38.7	47.8
6	123	4.6	35.3	33.1
7	184	6.8	28.8	44.1
8	192	7.1	27.6	53.9
9	164	6.1	44.2	40.1
10	178	6.6	37.7	37.6
11	247	9.1	24.6	75.9
12	157	5.8	26.4	66.3
13	202	7.5	22.9	76.5
14	179	6.6	27.1	73.7
15	173	6.4	30.7	78.5
Overall	2700	100.0	32.2	55.8

Table B16 Overall results of the test set for PNN II.1

PNN II.1				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	177	6.6	32.6	41.3
2	176	6.5	38.5	27.5
3	112	4.1	32.4	16.0
4	284	10.5	32.0	48.3
5	91	3.4	42.0	16.3
6	123	4.6	33.1	18.3
7	230	8.5	32.5	48.0
8	278	10.3	27.2	71.7
9	99	3.7	47.1	24.3
10	79	2.9	42.1	14.5
11	297	11.0	18.8	79.5
12	126	4.7	20.7	30.3
13	248	9.2	24.3	53.6
14	190	7.0	31.3	68.4
15	190	7.0	35.6	70.3
Overall	2700	100.0	30.8	42.3

Table B17 Overall results of the test set for PNN II.2

PNN II.2				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	216	8.0	32.6	55.1
2	126	4.7	35.4	36.3
3	205	7.6	28.6	42.3
4	215	8.0	27.8	46.7
5	254	9.4	40.7	49.4
6	75	2.8	31.8	13.1
7	243	9.0	27.3	52.5
8	240	8.9	26.3	59.2
9	48	1.8	47.7	8.6
10	61	2.3	30.6	10.2
11	292	10.8	21.7	70.8
12	176	6.5	28.0	36.5
13	231	8.6	20.2	52.5
14	184	6.8	28.8	67.3
15	134	5.0	27.8	55.8
Overall	2700	100.0	28.9	44.2

Table B18 Overall results of the test set for PNN II.3

PNN II.3				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	174	6.4	34.4	52.1
2	135	5.0	34.5	42.0
3	238	8.8	29.3	53.1
4	217	8.0	29.2	46.1
5	219	8.1	40.7	47.8
6	103	3.8	31.9	17.1
7	214	7.9	29.1	47.1
8	226	8.4	25.4	59.2
9	68	2.5	46.3	18.4
10	91	3.4	30.0	19.9
11	250	9.3	23.2	65.1
12	193	7.1	23.4	48.9
13	209	7.7	23.9	55.7
14	233	8.6	30.0	80.1
15	130	4.8	50.7	33.1
Overall	2700	100.0	30.7	46.0

Table B19 Overall results of the test set for PNN II.4

PNN II.4				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	162	6.0	31.8	51.5
2	186	6.9	38.4	43.5
3	172	6.4	28.2	50.9
4	238	8.8	33.6	48.3
5	216	8.0	36.7	43.8
6	92	3.4	37.8	22.3
7	195	7.2	31.8	38.7
8	262	9.7	29.8	64.9
9	79	2.9	44.6	20.4
10	100	3.7	40.3	19.4
11	279	10.3	24.1	64.1
12	245	9.1	24.8	60.7
13	177	6.6	25.7	42.6
14	129	4.8	28.3	37.4
15	168	6.2	50.9	40.7
Overall	2700	100.0	32.5	43.6

Table B20 Overall results of the test set for PNN II.5

PNN II.5				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	232	8.6	33.5	65.9
2	121	4.5	38.3	20.7
3	255	9.4	34.3	49.1
4	204	7.6	34.5	26.1
5	137	5.1	42.5	25.8
6	120	4.4	44.9	18.3
7	173	6.4	33.6	30.4
8	218	8.1	29.1	33.5
9	132	4.9	50.1	43.4
10	81	3.0	38.3	11.8
11	391	14.5	26.1	63.1
12	33	1.2	26.0	3.9
13	298	11.0	23.5	64.5
14	151	5.6	27.9	52.6
15	154	5.7	27.6	65.1
Overall	2700	100.0	32.5	38.0

Table B21 Overall results of the test set for PNN II.6

PNN II.6				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	243	9.0	36.5	66.5
2	130	4.8	37.6	21.2
3	246	9.1	33.4	46.9
4	131	4.9	37.6	17.2
5	154	5.7	43.4	25.8
6	54	2.0	34.3	9.7
7	118	4.4	33.7	18.6
8	87	3.2	34.1	9.9
9	53	2.0	44.9	21.7
10	64	2.4	40.1	11.3
11	489	18.1	28.5	72.3
12	150	5.6	46.2	16.9
13	353	13.1	33.4	57.4
14	189	7.0	34.6	66.7
15	239	8.9	40.0	83.1
Overall	2700	100.0	35.6	36.0

Table B22 Overall results of the test set for PNN II.7

PNN II.7				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	225	8.3	42.5	29.9
2	48	1.8	40.8	5.7
3	112	4.1	46.8	16.0
4	102	3.8	35.4	9.4
5	146	5.4	51.7	17.4
6	23	0.9	60.1	1.1
7	238	8.8	47.4	21.6
8	454	16.8	44.1	46.1
9	215	8.0	50.2	38.8
10	59	2.2	55.6	8.1
11	450	16.7	50.4	50.8
12	134	5.0	30.0	34.8
13	156	5.8	32.0	29.0
14	162	6.0	30.2	64.9
15	176	6.5	36.6	69.8
Overall	2700	100.0	43.6	29.3

Table B23 Overall results of the test set for PNN II.8

PNN II.8				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	286	10.6	38.9	64.7
2	110	4.1	43.4	17.1
3	193	7.1	32.4	38.3
4	167	6.2	41.2	21.1
5	208	7.7	40.9	23.0
6	99	3.7	36.1	14.3
7	184	6.8	31.0	27.9
8	171	6.3	31.0	27.7
9	139	5.1	43.5	14.5
10	130	4.8	36.1	16.1
11	310	11.5	26.8	65.6
12	177	6.6	25.7	53.9
13	208	7.7	25.6	67.8
14	153	5.7	26.7	60.2
15	165	6.1	28.7	67.4
Overall	2700	100.0	33.3	38.6

Table B24 Overall results of the test set for PNN II.9

PNN II.9				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	267	9.9	35.6	76.6
2	137	5.1	42.5	24.4
3	168	6.2	34.7	44.6
4	120	4.4	36.1	25.0
5	216	8.0	39.9	41.0
6	119	4.4	36.6	19.4
7	165	6.1	31.1	27.0
8	321	11.9	30.3	53.4
9	112	4.1	43.1	27.6
10	178	6.6	39.9	28.0
11	211	7.8	28.9	43.1
12	145	5.4	25.0	43.8
13	226	8.4	24.3	58.5
14	154	5.7	27.1	64.9
15	161	6.0	30.3	72.7
Overall	2700	100.0	33.2	43.0

Table B25 Overall results of the test set for PNN II.10

PNN II.10				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	271	10.0	34.9	71.3
2	112	4.1	42.7	19.2
3	138	5.1	32.4	41.1
4	98	3.6	42.1	14.4
5	235	8.7	43.4	36.0
6	62	2.3	40.7	10.9
7	124	4.6	34.0	22.1
8	367	13.6	33.9	51.3
9	128	4.7	49.1	28.9
10	166	6.1	42.2	21.5
11	278	10.3	30.5	51.8
12	130	4.8	29.9	37.6
13	232	8.6	27.2	54.6
14	149	5.5	27.0	50.3
15	210	7.8	47.1	44.2
Overall	2700	100.0	36.3	36.8

Table B26 Overall results of the test set for PNN I.1

PNN I.1				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	0	0.0	0.0	0.0
2	0	0.0	0.0	0.0
3	0	0.0	0.0	0.0
4	0	0.0	0.0	0.0
5	40	1.5	62.8	3.4
6	0	0.0	0.0	0.0
7	86	3.2	51.1	5.4
8	242	9.0	46.3	16.8
9	114	4.2	47.3	26.3
10	0	0.0	0.0	0.0
11	938	34.7	60.3	79.5
12	51	1.9	68.1	3.4
13	657	24.3	69.0	44.8
14	323	12.0	60.8	57.3
15	249	9.2	49.7	69.2
Overall	2700	100.0	59.6	20.3

Table B27 Overall results of the test set for PNN I.2

PNN I.2				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	180	6.7	33.5	41.9
2	24	0.9	58.0	6.2
3	321	11.9	31.1	44.6
4	421	15.6	35.8	55.0
5	0	0.0	0.0	0.0
6	76	2.8	35.1	11.4
7	300	11.1	38.2	45.6
8	348	12.9	33.7	64.4
9	0	0.0	0.0	0.0
10	0	0.0	0.0	0.0
11	555	20.6	27.6	83.1
12	195	7.2	24.1	38.2
13	124	4.6	27.4	24.0
14	156	5.8	32.9	40.4
15	0	0.0	0.0	0.0
Overall	2700	100.0	32.2	31.0

Table B28 Overall results of the test set for PNN I.3

PNN I.3				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	431	16.0	37.5	79.0
2	80	3.0	45.4	8.3
3	113	4.2	34.4	12.6
4	213	7.9	39.5	20.0
5	216	8.0	42.6	17.4
6	0	0.0	0.0	0.0
7	133	4.9	35.9	11.8
8	304	11.3	33.6	26.7
9	66	2.4	38.4	4.6
10	0	0.0	0.0	0.0
11	478	17.7	29.8	51.8
12	0	0.0	0.0	0.0
13	373	13.8	26.7	62.3
14	82	3.0	24.5	18.1
15	211	7.8	29.4	61.6
Overall	2700	100.0	33.8	24.9

Table B29 Overall results of the test set for PNN I.4

PNN I.4				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	395	14.6	36.7	78.4
2	0	0.0	0.0	0.0
3	154	5.7	35.8	23.4
4	0	0.0	0.0	0.0
5	148	5.5	43.7	14.6
6	0	0.0	0.0	0.0
7	104	3.9	39.0	11.3
8	185	6.9	38.2	17.3
9	0	0.0	0.0	0.0
10	2	0.1	40.7	0.5
11	378	14.0	32.7	40.5
12	132	4.9	44.5	6.7
13	463	17.1	32.3	63.4
14	254	9.4	68.6	34.5
15	485	18.0	67.5	52.3
Overall	2700	100.0	44.8	22.6

Table B30 Overall results of the test set for PNN I.5

PNN I.5				
Sky	No. of occurrence	Frequency of occurrence (%)	RMSE (%)	Hit percentage (%)
1	296	11.0	58.4	29.3
2	0	0.0	0.0	0.0
3	228	8.4	49.9	28.6
4	2	0.1	32.0	0.0
5	24	0.9	52.8	5.1
6	18	0.7	41.0	1.7
7	161	6.0	46.3	7.4
8	957	35.4	50.9	56.5
9	113	4.2	54.3	15.8
10	4	0.1	50.0	0.0
11	0	0.0	0.0	0.0
12	32	1.2	35.3	4.5
13	0	0.0	0.0	0.0
14	406	15.0	32.1	61.4
15	459	17.0	73.0	37.2
Overall	2700	100.0	52.2	16.1

APPENDIX C

LIST OF PUBLICATIONS

Journal

i. Li DHW and Tang HL (2008). Standard skies classification in Hong Kong. Journal of Atmospheric and Solar-Terrestrial Physics 70(8-9):1222-1230.

ii. Wan KKW, Tang HL, Yang L and Lam JC (2008). An analysis of thermal and solar zone radiation models using Angstrom-Prescott equation and artificial neural networks. Energy 33(7):1115-1127.

iii. Lam JC, Tang HL and Li DHW (2008). Seasonal variations in residential and commercial sector electricity consumption in Hong Kong. Energy 33(3):513-523.

iv. Li DHW, Lam TNT, Cheung KL and Tang HL (2008). An analysis of luminous efficacies under the CIE standard skies. Renewable Energy 33(11):2357-2365.

Conference paper

i. Li DHW, Tang HL, Wong SL, Tsang EKW, Cheung GHW and Lam TNT (2007). Skies classification using artificial neural networks (ANN) techniques. The 6th International Conference on Indoor Air-Quality, Ventilation & Energy Conservation in Buildings, Sendai. Proceeding I, p.61-68.

ii. Li DHW and Tang HL (2008). A study of artificial neural networks to classify the 15 CIE standard skies. International Conference on Intelligent Systems, Structure and Facilities, Hong Kong. p.12-21.