MULTIOBJECTIVE VISIBLE SPECTRUM OPTIMIZATION: A GENETIC ALGORITHM APPROACH

by

Neil Holger White Eklund

A Thesis Submitted to the Graduate
Faculty of Rensselaer Polytechnic Institute
in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
Major Subject: Engineering Science

Approved by the
Examining Committee:

_____
Mark J. Embrechts, Thesis Advisor

_____
Peter R. Boyce, Member

_____
Curt M. Breneman, Member

_____
Ralph G. Noble, Member

_____
John Van Derlofske, Member

Rensselaer Polytechnic Institute
Troy, New York

September 2002
(for graduation December 2002)

© Copyright 2002
by
Neil Holger White Eklund
All Rights Reserved

Table of Contents

Table of Contents iii

List of Tables vi

List of Figures vii

Dedication ix

Acknowledgements x

Abstract xi

1. Introduction 1

1.1 Previous Work on Spectrum Optimization 3

1.2 Dissertation Outline 7

2. Color and the Measurement of Color 10

2.1 The CIE Method for Color Specification 11

2.2 Color Rendering 14

2.2 Color Temperature and Correlated Color Temperature 17

3. Genetic Algorithms and Multiobjective Optimization 20

3.1 Genetic Algorithms 21

3.2 Multiobjective Optimization 25

3.3 GAs for Multiobjective Optimization 27

3.3.1 Approaches Employing Aggregating Functions 28

3.3.2 Non-Pareto Based Approaches 30

3.3.3 Pareto Based Approaches 32

4. Optimizing for Luminous Efficacy and Chromaticity Only 34

4.1 Genetic Algorithm Implementation 35

4.1.1 Chromosome Encoding 36
4.1.2 Evaluating Fitness 37
4.1.3 Selection 39
4.1.4 Crossover 40
4.1.5 Mutation 41
4.2 Efficacy & Chromaticity Results 45
4.2.1 The Pareto-Optimal Chromaticity-Efficiency Surface 51
4.3 Implications for Future Work 55
5. Main Results: Optimizing for Efficacy, Chromaticity, and CRI 57
5.1 Target Objectives Genetic Algorithm 58
5.2 TOGA Implementation 61
5.2.1 Chromosome Encoding 64
5.2.2 Evaluating Fitness 68
5.2.3 Selection and Crossover 68
5.2.4 Mutation 69
5.3 Efficacy, Chromaticity & CRI Results 71
5.3.1 GA Results 71
5.3.2 Lighting Results 79
6. Conclusions and Future Work 87
6.1 Practical Application of this Method 90
6.2 Future Work 92
7 References 94
Appendix A: Additional Figures 107

Appendix B: Matlab Code 116

Appendix C: Related Papers 129

List of Tables

Table 1. Selected milestones in lighting technology (IESNA 2000). 1

Table 2. Chromaticity coordinates and efficiency of the filtered spectra. 46

Table 3. Target values for CRI and chromaticity, with corresponding color temperature. 62

Table 4. Percentage of filtered spectrum chromaticity coordinates falling within MacAdam ellipses of different size. 72

Table 5. Percentage of filtered spectra with CRI within one to five units of the target CRI. 74

Table 6. Range of values encompassing the center 50% (i.e., the interquartile range), 90%, and 98% of the variability in efficiency. 76

Table 7. Matlab files listed, and a brief description of the function of each. 116

List of Figures

Figure 1. Luminous efficacy of common electric light sources (and the sulphur lamp). ... 3
Figure 2. CIE color matching functions (dimensionless) for the 1931 CIE Standard Colorimetric Observer. ... 11
Figure 3. 1931 CIE (*x,y*) chromaticity diagram. ... 13
Figure 4. Reflectance of the CIE standard CRI test samples. ... 17
Figure 5. Chromaticity diagram with blackbody locus ... 19
Figure 6. The GA optimization process. ... 23
Figure 7. Concepts of multiobjective optimality. ... 27
Figure 8. SPDs for Metal Halide, High Pressure Sodium and Incandescent lamps. ... 35
Figure 9. An poor solution for the incandescent lamp. ... 38
Figure 10. A good solution for the incandescent lamp (cf., Figure 9). ... 39
Figure 11. Probability of selection for each rank. ... 40
Figure 12. A sub-optimal filter. ... 43
Figure 13. BCM applied to part of region A (bold line). ... 43
Figure 14. PM applied to part of region C (bold line). ... 44
Figure 15. SM applied to part of region B (bold line). ... 45
Figure 16. Filter transmittance for the three spectra ... 47
Figure 17. Path of the fittest chromosome through chromaticity diagram. ... 49
Figure 18. Path of the fittest chromosome through the search space ... 50
Figure 19. The points sampled for determining the chromaticity-efficiency surface. ... 52
Figure 20. Efficiency contours for the incandescent source. ... 52
Figure 21. Efficiency contours for the MH source. ... 53
Figure 22. Efficiency contours for the HPS source. ... 53
Figure 23. Normalized luminous efficacy and chromatic efficacy. ... 55
Figure 24. SPDs for incandescent, high pressure sodium, metal halide, and sulphur lamps. ... 63
Figure 25. Positions of unfiltered spectra and target chromaticities on the CIE 1931 chromaticity diagram. ... 63
Figure 26. Boxed region of Figure 25. ... 64
Figure 27. Trapezoid encoding. ... 65
Figure 28. Intermediate steps to decoding a chromosome. ... 67

Figure 29. Standard deviation for normal mutation as a function of generation. 70

Figure 30. Distribution of chromaticity coordinate error by color temperature and CRI, incandescent spectrum. 73

Figure 31. Boxplots of CRI error by color temperature and CRI, for all spectra (S, sulphur; I, incandescent; H, high pressure sodium; M, metal halide)..... 75

Figure 32. Boxplots of efficiency by color temperature and CRI, for all spectra (S, sulphur; I, incandescent; H, high pressure sodium; M, metal halide)..... 77

Figure 33. Relationship between efficiency and CRI at two different color temperatures. 79

Figure 34. Median efficiency of filtered spectra for the four different lamp types. For each set of axes, the ordinate is efficiency and the abscissa is CRI..... 81

Figure 35. Tradeoff between efficiency and CRI at each chromaticity..... 83

Figure 36. Median (solid line), 25th percentile (dashed line), and 75th percentile (dotted line) of the transmittance (at each wavelength sampled) of the filter designed for the incandescent spectrum, at each $\overline{T}^c$ 86

Figure 37. Distribution of chromaticity coordinate error by color temperature and CRI, sulphur lamp spectrum..... 108

Figure 38. Distribution of chromaticity coordinate error by color temperature and CRI, HPS spectrum..... 109

Figure 39. Distribution of chromaticity coordinate error by color temperature and CRI, MH spectrum. 110

Figure 40. Median (solid line), 25th percentile (dashed line), and 75th percentile (dotted line) of the transmittance (at each wavelength sampled) of the filter designed for the sulphur lamp spectrum, at each $\overline{T}^c$ 111

Figure 41. Median (solid line), 25th percentile (dashed line), and 75th percentile (dotted line) of the transmittance (at each wavelength sampled) of the filter designed for the MH spectrum, at each $\overline{T}^c$ 112

Figure 42. Median (solid line), 25th percentile (dashed line), and 75th percentile (dotted line) of the transmittance (at each wavelength sampled) of the filter designed for the HPS spectrum, at each $\overline{T}^c$ 113

Figure 43. Efficiency contours for the incandescent source plotted in the CIE 1976 (u',v') coordinate system..... 114

Figure 44. Efficiency contours for the MH source plotted in the CIE 1976 (u',v') coordinate system..... 115

Figure 45. Efficiency contours for the HPS source plotted in the CIE 1976 (u',v') coordinate system..... 115

For the Cammer girls: Jan, Diane, and Evelyn.

Acknowledgements

I would like to extend my deepest thanks to the following five people. Without the various contributions of each of them, I would not be where I am today!

First and foremost, my mother and grandmother for their love and support of me throughout my life.

Second, my wonderful wife Julie, who has been incredibly supportive of me (whether I deserved it or not) as long as I have known her.

Third, my good friend and mentor, Dr. Peter Boyce, who has (once again) provided thoughtful comments and careful editing. I hope I can be as kind to my students as he has been to me.

Finally, Dr. Mark Embrechts who revolutionized my research interests by introducing me to soft computing. I am grateful for his invaluable help and guidance on my research, particularly on my dissertation.

I would also like to thank Dr. Curt Breneman, Dr. Ralph Noble, and Dr. John Van Derlofske for taking the time to serve on my dissertation committee, and for their helpful comments.

Abstract

One of the principal goals motivating advances in lighting technology is maximization of luminous efficacy, the ratio of the total luminous flux to total power input (i.e., the "amount of light" per Watt). However, colorimetric properties, such as the apparent color and color rendering properties of the light, have a strong influence on the application and adoption of new light source technologies, probably because these properties can easily be directly assessed by consumers. There are cases where the efficacy of a new electric light source technology is very high but the color properties are unsuitable.

It is possible to exchange some efficacy for better colorimetric properties: there are an infinite number of ways to filter a broad-spectrum light so that it has better colorimetric properties. However, almost all of these filters will reduce luminous efficacy by an unacceptable amount. A novel approach to multiobjective optimization, the *target objectives genetic algorithm* (TOGA), is introduced, and implemented to determine the tradeoff between luminous efficacy, chromaticity, and color rendering for four filtered lamp spectra.

TOGA is a non-Pareto, non-aggregating function approach to multiobjective optimization similar in concept to goal programming. TOGA is computationally very fast, generates multiple optimal points during each run, and can be used for any multiobjective optimization problem. However, TOGA is most efficient when the researcher has some domain knowledge and is able to select good combinations of objectives. In addition, like

goal programming, TOGA can result in solutions that are not Pareto-optimal (but are optimal solutions at a particular level of target objectives).

The resulting filtered lamp spectra are examined with respect to GA performance, and the deviation from the target chromaticities and target color rendering values are shown to be largely well within lighting industry limits, and the variance in efficiency of the resulting filtered lamp spectra is shown to be low. This technique can be use to determine how new electric light source technologies might be filtered so that they are more commercially viable.

1. Introduction

The research on more efficient electrical light source technology is both well funded and highly motivated. Annual sales (worldwide) of electrical lighting is estimated to be over $15 billion (COST 2001). According to COST (European Co-operation in the field of Scientific and Technical Research), "currently, more than 7.5 billion lamps operate world-wide consuming 1,000 billion kWh per year (10-15% of the global energy production world-wide)" (COST 2001). Thus, periodically new electric light source technologies are developed, or major improvements on old technologies are made (Table 1). Moreover, "…the industry is dynamic, and the introduction of new light sources is expected to continue at least at the present rate well into the next century." (IESNA 2000; e.g., Gu, Burrows, and Venkatesh, 1997).

Table 1. Selected milestones in lighting technology (IESNA 2000).

Year	Milestone
1810	electric arc lamp demonstrated
1840	early electric incandescent lamp demonstrated
1862	electric arc lamp in use
1878	Swan (England) and Edison (United States) independently develop practical filament incandescent lamp
1904	discharge lamps developed
1907	tungsten filament developed
1910	neon lamp developed
1913	inert gas filling for incandescent bulbs developed
1931-33	discharge lamps developed
1938	fluorescent lamp developed
1959	tungsten halogen lamp developed
1961	metal halide lamps developed
1962	first practical visible light emitting diode developed
1967-68	high pressure sodium lamp developed
1980-84	compact ("folded") fluorescent developed
1990	sulphur lamp developed
1992	induction lamp developed
1993	indium-gallium-nitride light emitting diodes developed

One of the principal goals motivating advances in lighting technology is maximization of luminous efficacy, the ratio of the total luminous flux to total power input (i.e., the "amount of light" per Watt; Wyszecki and Stiles, 1982). However, there are many factors that affect the commercial viability of a new light source technology, including colorimetric properties (apparent color[1], color rendering), lamp life, luminous efficacy and color stability when dimmed, source size, warm up time, color uniformity, color stability over life of the source, and lumen maintenance (light output over the life of the source). Colorimetric properties have a strong influence on the application and adoption of new light source technologies, probably because an observer (e.g., a consumer or client) can easily assess them (cf., lumen maintenance). For example, even though metal halide lamps (which produce a white light that renders color well) are less efficient and more expensive to manufacture than low pressure sodium lamps (which produce a yellow light that renders color poorly), there are many more applications in practice for metal halide lamps because they have much better colorimetric properties.

Ideally, a new electric light source technology will have high luminous efficacy, have acceptable apparent color (typically “white”), and render colors well. While getting all three properties in a new light source technology just right is rare, there are cases where efficacy is very high and the color properties are just slightly off. For example, the sulphur lamp (Siminovitch, Gould, and Page, 1997; Turner, et al., 1997) has very high luminous efficacy (Figure 1), a color rendering index (CRI) of 78, and a greenish-white apparent color (1931 CIE chromaticity coordinates x=.33, y=.41, a poor color for many

[1] The color of light coming directly from the source (cf., reflected light).

applications). It is possible to exchange some efficacy for better colorimetric properties: there are an infinite number of ways to filter a broad-spectrum light so that it has better colorimetric properties. However, almost all of these filters will reduce luminous efficacy by an unacceptable amount.

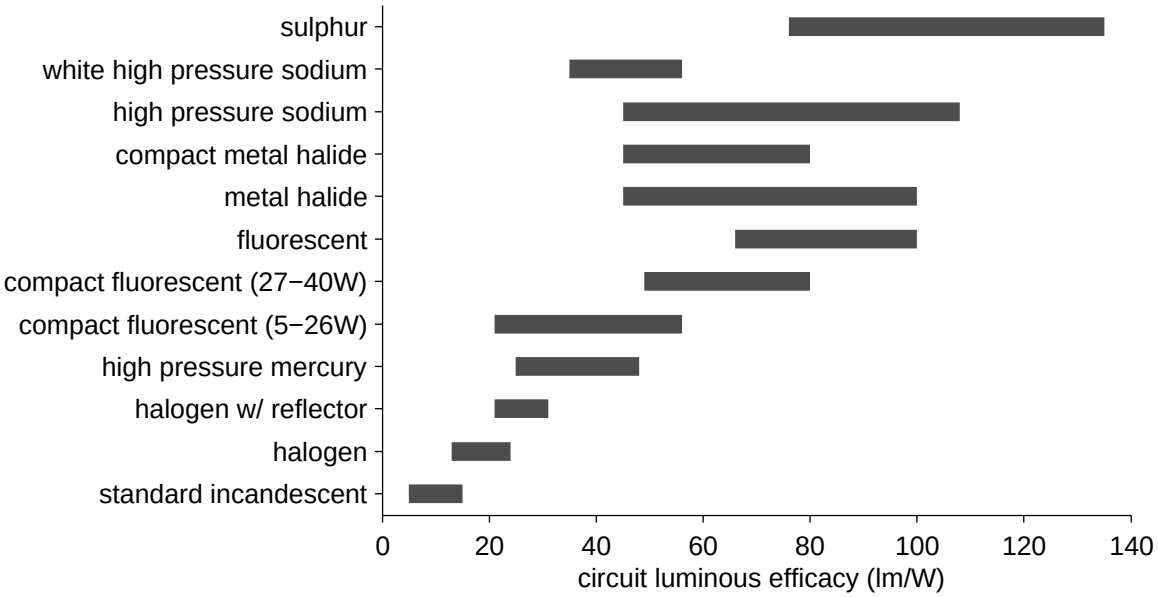

Figure 1. Luminous efficacy of common electric light sources (and the sulphur lamp).

This dissertation is concerned with the tradeoff between luminous efficacy, apparent color, and color rendering properties of light sources. In particular, it describes a method for determining this tradeoff using a novel approach to multiobjective optimization employing genetic algorithms.

1.1 Previous Work on Spectrum Optimization

There are several different perspectives from which to examine the idea of spectrum optimization, depending on one's application. One may wish to optimize a spectrum such that (from least to most complex) it:

- maximizes luminous efficacy, regardless of color

- maximizes luminous efficacy for a particular chromaticity
- maximizes luminous efficacy for a particular CRI
- maximizes luminous efficacy for a particular chromaticity and CRI
    - for line or band spectra[2]
    - for continuous spectra[3]

The history of spectrum optimization follows this progression, principally because of advances in optimization theory and computing power rather than lack of interest in more complex problems.

In 1924, the CIE[4] adopted a standard (based on psychophysical measurements of brightness matching) for photopic spectral luminous efficiency (which peaks at 555 nm) and a factor to convert radiant flux to photopic luminous flux, 683 lm/W (Wyszecki and Stiles, 1982). Therefore, by definition, the maximum luminous efficacy, where all of the electrical power is converted into a spectral line at 555nm, is 683 lm/W. However, this (theoretical) lamp, which optimizes efficacy alone, has terrible color properties: all objects appear yellowish green, gray, or black.

MacAdam (1950) calculated the maximum attainable luminous efficacy at all chromaticities, which (for all non-spectral chromaticities) require the mixture of a single pair of wavelengths only. However, a source of light designed on this principal would

---

[2] Line spectra have energy at discrete wavelengths, with very narrow bandwidth (e.g., <1 nm), and no energy elsewhere; band spectra have energy at discrete wavelengths, with wider bandwidth (e.g., 10 nm), and no energy elsewhere.
[3] Continuous spectra have energy at all wavelengths of the visible spectrum, and no sharp transitions (i.e., they are smooth).
[4] Commission Internationale de l'Eclairage

have poor color rendering, and "…would not be a satisfactory source of artificial daylight illumination." (MacAdam, 1950). However, this represents an absolute upper bound against which the efficacy of real sources can be compared. Ivey (1963) expanded on MacAdam's work, taking into account the emission spectra of real phosphors[5] and the dependence of energy in a photon on frequency (e.g., assuming equal quantum efficiency, a 700 nm monochromatic phosphor will emit only 57% of the power input compared to a 400 nm monochromatic phosphor). Ivey's bounds were similar in shape to MacAdam's (1950), although substantially reduced in magnitude.

MacAdam (1935a) presents a proof of a theorem that allows the optimal spectral reflectance for a pigment to achieve a maximum colorimetric purity for a given illuminant and dominant wavelength to be determined. This can also be used (MacAdam, 1935a, 1935b) to determine how an arbitrary SPD may be filtered to achieve any chromaticity at maximum efficiency. However, this method offers no guarantee that the color rendering will be acceptable (and it is likely to be poor for many chromaticities).

Until this point, work had been done on optimizing efficiency at a given chromaticity. However, this approach is insufficient for the needs of the lighting industry: it is also necessary to account for color rendering properties of light.

Koedam and Opstelten (1971) use trial and error (guided by color theory) to develop three-line spectra on the blackbody locus with CRI ≅ 80. Koedam, Opstelten, and

[5] Which, "…for most phosphors … consist of a single broad bell-shaped curve, the width and peak wavelength depending on the particular phosphor considered."

Radielović (1972) examined (again via trial and error) the effect of using three bands of differing bandwidth on color rendering and efficacy, although they only look at a few points. Thornton (1971) uses a similar approach to explore the tradeoff between CRI and efficiency of some three line spectra (for white light). All three of these papers identify similar regions of the spectrum (around 450, 540, and 610 nm) as being particularly important for color rendering in line or band spectra.

Einhorn and Einhorn (1967), Walter (1971), Haft and Thornton (1972), and Opstelten, Radielović, and Verstegen (1975) were the first researchers to examine in a systematic way the relationship between CRI and efficacy for certain chromaticities, using three line or three band spectra. All four papers present calculations (starting with different assumptions) of CRI/efficacy Pareto optimal fronts for different colors, and note that the lighting industry was manufacturing many lamps that were far from this front (i.e., even given physical limitations, there was substantial room for improvement).

Walter (1978) applied nonlinear programming to the spectrum optimization problem for efficacy and color rendering at a particular chromaticity. Although he was never able to get the algorithm to converge, he was able to find three and four line spectra with high efficiency and high CRI.

Other researchers have applied a variety of approaches to solve related problems. Ohta and Wyszecki (1976b) use linear programming to design illuminants that render a limited number of objects at desired chromaticities. Ohta and Wyszecki (1976a) use nonlinear

programming to explore the relationship between illuminant changes and color change (in relation to setting tolerances for illuminant differences). Dupont (2002) applied a variety of methods (including neural networks, the simplex method, genetic algorithms, simulated annealing, et al.) to reconstruct reflectance curves from tristimulus values for a given SPD. Denbigh and Jones (1984) developed software to aid in the development of high pressure sodium lamps, which among other things incorporated previous research on spectrum optimization to propose lamp parameters that were likely to result in lamps with high CRI. Numerous papers describe experiments in lamp spectrum optimization (e.g., Krasko, Brates, and Nortrup, 1998; Carleton, Seinen, and Stoffels, 1997) dealing directly with the physical properties of the lamp (e.g., voltage, temperature, pressure, wall thickness, chemical components, etc.).

1.2 Dissertation Outline

The remainder of this document is arranged in six chapters and three appendices. Chapter 2 defines color, and describes a standard method (used throughout industry since the 1930s) for its measurement and quantification. The process for calculating a lighting industry standard for the measurement of color rendering (the Color Rendering Index, CRI) is also described. Finally, the concept of color temperature (a one-dimensional measure of chromaticity), which is used by the lighting industry to specify lamp apparent color, is introduced.

Chapter 3 provides an introduction to the basic mechanics of genetic algorithms, and introduces the concept of Pareto-optimality in multiobjective optimization. Three broad approaches for using genetic algorithms for multiobjective optimization (aggregation

methods, non-Pareto methods, and Pareto methods) are described, and examples are given of each.

Chapter 4 describes a genetic algorithm (GA) technique for the maximization of the filtered luminous efficacy of an arbitrary spectral power distribution (e.g., based on a new lamp technology) at a particular chromaticity. A chromosome encoding method and three related problem-specific mutation methods are presented. The results from the GA technique are shown to compare favorably to known optima, particularly at areas of likely interest for industrial applications, suggesting that the GA approach is suitable to this type of problem. However, some serious limitations of this approach are discussed (relating particularly to computational efficiency) and a direction for further research is identified.

Chapter 5 presents the main results of this work. A novel approach to using genetic algorithms for multiobjective optimization, the Target Objectives Genetic Algorithm (TOGA), is described, and advantages and disadvantages of this technique are discussed. TOGA is implemented to determine the tradeoff between luminous efficacy, chromaticity, and color rendering for four filtered lamp spectra. The results are examined with respect to GA performance, and the deviation from the target chromaticities and target color rendering values are shown to be well within lighting industry limits, and the variance in efficiency of the resulting filtered lamp spectra is shown to be low. The results are also examined from a lighting perspective, and, for the range of color rendering and chromaticity examined, while there was a clear interaction, chromaticity

has a much greater effect on the maximum attainable luminous efficacy than color rendering. TOGA is also shown to produce filters with efficiencies quite near the maximum possible for any filtered light source at a particular chromaticity (i.e., regardless of color rendering).

The main conclusions and suggestions for future work are discussed in Chapter 6. Fundamental results, in relation to both genetic algorithms and lighting, are presented, along with a description of the practical application of this method. The advantages and limitations of the TOGA method are discussed, and recommendations for future work are made.

Appendix A presents some additional figures relating to lighting. Appendix B provides the code used for the results presented in Chapter 5 (the TOGA implementation). This is included for archival purposes only, rather than as ready to use software. Appendix C contains several related papers by this author that have been published elsewhere.

2. Color and the Measurement of Color

> *All vision is colour vision, for it is only by observing differences of colour that we distinguish the forms of objects. I include differences of brightness among differences of colour*.
>
> -James Clerk Maxwell, 1872

Color is the aspect of electromagnetic radiation evaluated by the visual system that allows an observer to distinguish two homogeneous fields of the same size and shape that differ only in spectral composition. Differences in spectra may elicit differences in the perception of hue, brightness or saturation. Hue is the attribute of visual sensation whereby an area might be described as either purely, or some combination of, red, green, yellow, or blue (this is the aspect that laymen usually think of as "color"). Brightness is the attribute of visual sensation whereby an area appears to emit more or less light (at one extreme barely perceptible, at the other dazzling). Saturation is the attribute of visual sensation describing the amount of whiteness in a particular hue (e.g., although they are of the same hue, pink is less saturated than red).

Colorimetry is the branch of physics concerned with the quantification and prediction of color perception based on the radiometric properties of a stimulus. Colorimetry can be used to calculate various properties of a given spectral power distribution (SPD, the radiant power per unit wavelength as a function of wavelength) relating to how that SPD will appear to a typical observer with normal color vision. Colorimetry is used extensively in industry (Judd and Wyszecki, 1975), e.g., to specify colors of materials or to specify tolerances for color variation across different lots (Alman, 1993; CIE 1995a).

This chapter describes three metrics that are commonly used to describe light sources: chromaticity coordinates, the color rendering index, and (correlated) color temperature.

2.1 The CIE Method for Color Specification

The Commission Internationale de l'Eclairage (CIE) is the International Standards Organization recognized body for all matters regarding the science and art of lighting. Although each person perceives color slightly differently, the CIE has developed a colorimetric system to quantify the color of a light source from its SPD based on a "standard observer" (Judd, 1933). The properties of the standard observer are based on psychophysical measurements of color matching (Guild, 1931; Wright, 1928-29), and are used to derive the color matching functions (Fairman, Brill, and Hemmendinger, 1997; Wright, 1981b). The CIE 1931 color matching functions (Figure 2) are fundamental to colorimetry, and can be used to define a coordinate system for determining if two different spectra (when seen under identical conditions) will match or appear different (Kaiser and Boynton, 1996).

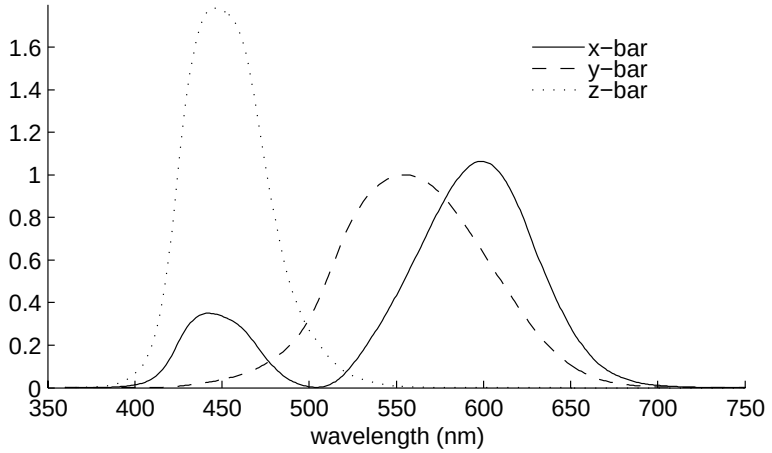

Figure 2. CIE color matching functions (dimensionless) for the 1931 CIE Standard Colorimetric Observer. These functions are based on psychophysical measurements of color matching, and can be used to define a coordinate system for determining if two different spectra (when seen under identical conditions) will match or appear different.

The X, Y and Z tristimulus values are the amount of three imaginary primaries required to produce a color, with the Y tristimulus value proportionate to the brightness of the color. Tristimulus values are calculated by weighting the SPD by the color matching functions and integrating over the visible spectrum (CIE, 1971):

$$X = k \int_{\lambda} P_{\lambda} \bar{x}(\lambda) d\lambda$$

$$Y = k \int_{\lambda} P_{\lambda} \bar{y}(\lambda) d\lambda$$

$$Z = k \int_{\lambda} P_{\lambda} \bar{z}(\lambda) d\lambda$$

where $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are the color matching functions, $P(\lambda)$ is the power at wavelength λ, and k is an application specific constant[6]. The unit plane, X + Y + Z = 1, of the tristimulus color space is known as the chromaticity diagram (Judd and Wyszecki, 1975), and points within this plane are known as x and y chromaticity coordinates:

$$x = \frac{X}{X + Y + Z}$$

$$y = \frac{Y}{X + Y + Z}$$

Figure 3 is a plot of the 1931 CIE chromaticity diagram. Monochromatic colors appear on the spectrum locus; monochromatic colors on the "purple line" are not possible – they are produced by a mixture of two (red and blue) primaries. All polychromatic colors have chromaticity coordinates within the bounds of the spectrum locus and purple line. White

[6] Unimportant for this application, and set to 1.0.

light is near the equal energy point, W, so colors near the equal energy point are less saturated than colors near the spectrum locus. Brightness is not encoded in the chromaticity diagram so, for example, broad spectrum white light reflected from an orange and from a chocolate bar will have similar chromaticity coordinates.

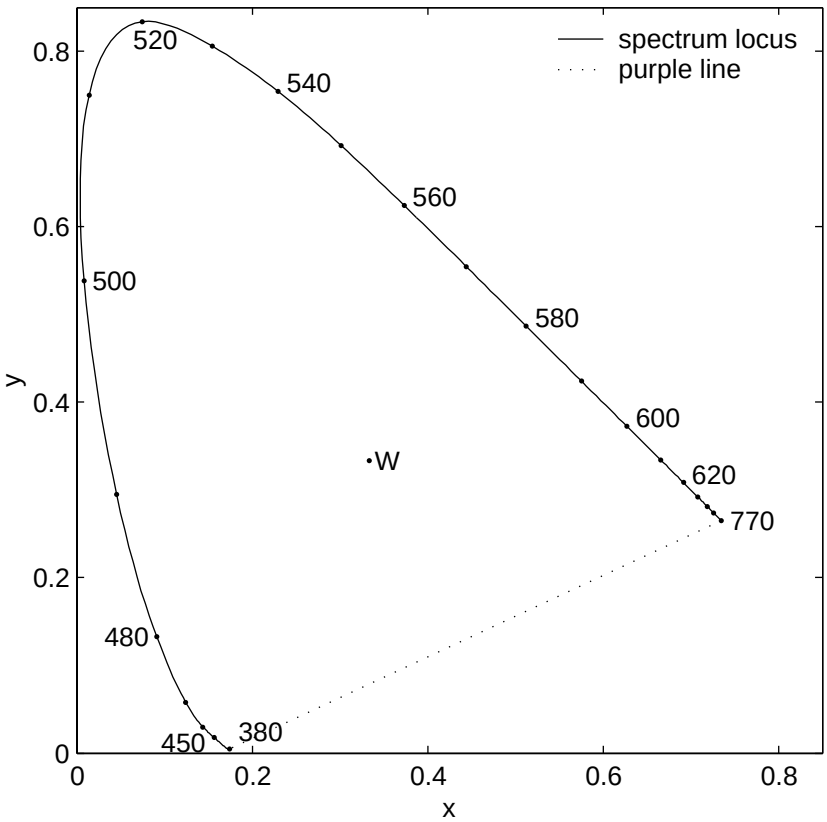

Figure 3. 1931 CIE (*x,y*) chromaticity diagram. Numbers are wavelength of indicated point (in nanometers); W is equal energy point. Monochromatic colors appear on the spectrum locus; colors on the "purple line" are produced by a mixture of two (red and blue) primaries.

There have been several proposals to update the 1931 CIE standard colorimetric observer (e.g., Judd, 1951; Vos, 1978; Shaw, 1997; Estevez, 1982; Stiles and Burch, 1955; Thornton, 1992b, 1992c, 1998). It is well known (Stockman and Sharpe, 1998; Thornton, 1992a; Vos, Estevez, and Walraven, 1990) that the luminous efficacy of the short wavelength (below 460 nm) region of the spectrum (particularly below 410 nm; Vos

1978; Smith and Pokorny, 1975) is underestimated in the CIE 1924 standard observer for luminance (which the 1931 standard colorimetric observer is partially based on), probably due to lenticular photoluminescence (Kandel, Eklund, and Schroeder, 1992; Kandel, et al., 1993; Said and Weale, 1959; Mäntyjärvi and Tuppurainen, 1996; Silk, 1999). However, despite this limitation, the 1931 CIE standard colorimetric observer has remained a useful international standard (Fairchild, 1998; Wright, 1981a), and it is used here.

2.2 Color Rendering

Color rendering refers to the degree to which a test illuminant elicits color appearance in reflected objects similar to a reference illuminant (typically daylight or a blackbody radiator of similar color temperature, although in principal, any reference may be used). The CIE (1995b) has defined the general color rendering index (CRI), which is a measure of the average appearance of eight standardized reflective color samples chosen to be of intermediate saturation and spread throughout the range of hues. Although CRI has several deficiencies as a metric (van Trigt, 1999), it has nonetheless become entrenched as a standard in the lighting industry, and no lamp is designed without giving thought to it.

The CIE method for determining color rendering uses a Planckian radiator as a standard for color temperatures below 5000 K; at or above 5000 K, one of a series of SPDs corresponding to phases of daylight are used (CIE, 1995b). CRI can range[7] from 0 to 100; a CRI of 100 indicates that the test colors appear exactly the same as under the reference

[7] Negative values of CRI can be calculated; however, these are generally considered invalid.

source, while lower values of CRI indicate that the colors appear progressively more different from the reference when illuminated by the test source. To calculate CRI by the CIE specified method (CIE, 1995b), colorimetric data (for both the test source and the reference) must be transformed from CIE 1931 system (x, y) to the (u, v) coordinates of the 1960 CIE Uniform Chromaticity Scale using:

$$u = 4X/(X + 15Y + 3Z)$$
$$v = 6Y/(X + 15Y + 3Z)$$

The adaptive color shift due to the different state of chromatic adaptation under the lamp to be tested, *k*, under the reference illuminant, *r*, for reflectance standard, *i*, is accounted for by:

$$u'_{k,i} = \frac{10.872 + 0.404\frac{c_r}{c_k}c_{k,i} - 4\frac{d_r}{d_k}d_{k,i}}{16.518 + 1.481\frac{c_r}{c_k}c_{k,i} - \frac{d_r}{d_k}d_{k,i}}$$

$$v'_{k,i} = \frac{5.520}{16.518 + 1.481\frac{c_r}{c_k}c_{k,i} - \frac{d_r}{d_k}d_{k,i}}$$

where:

$$c = (4 - u - 10v)/v$$
$$d = (1.708v + 0.404 - 1.481u)/v$$

These colorimetric data must now be transformed into the CIE 1964 Uniform Color Space coordinates by the following:

$$W^*_{r,i} = 25(Y_{r,i})^{1/3} - 17; \quad W^*_{k,i} = 25(Y_{k,i})^{1/3} - 17;$$
$$U^*_{r,i} = 13W^*_{r,i}(u_{r,i} - u_r); \quad U^*_{k,i} = 13W^*_{r,i}(u'_{k,i} - u'_k);$$
$$V^*_{r,i} = 13W^*_{r,i}(v_{r,i} - v_r); \quad V^*_{k,i} = 13W^*_{r,i}(v'_{k,i} - v'_k);$$

where the *Y* tristimulus values, $Y_{r,i}$ and $Y_{k,i}$, are normalized so that $Y_r = Y_k = 100$.

In this coordinate space, the color difference between the test color sample under the test lamp *k* and illuminated by the reference source *r* is calculated:

$$\Delta E = \sqrt{(\Delta U_i^*)^2 + (\Delta V_i^*)^2 + (\Delta W_i^*)^2}$$

The Special Color Rendering Index, R_i, for each test color sample is calculated by:

$$R_i = 100 - 4.6\Delta E_i$$

Finally, the general color rendering index is calculated by:

$$CRI = \frac{1}{8}\sum_{i=1}^{8} R_i$$

The reflectance of the eight test color samples specified by the CIE is plotted in Figure 4.

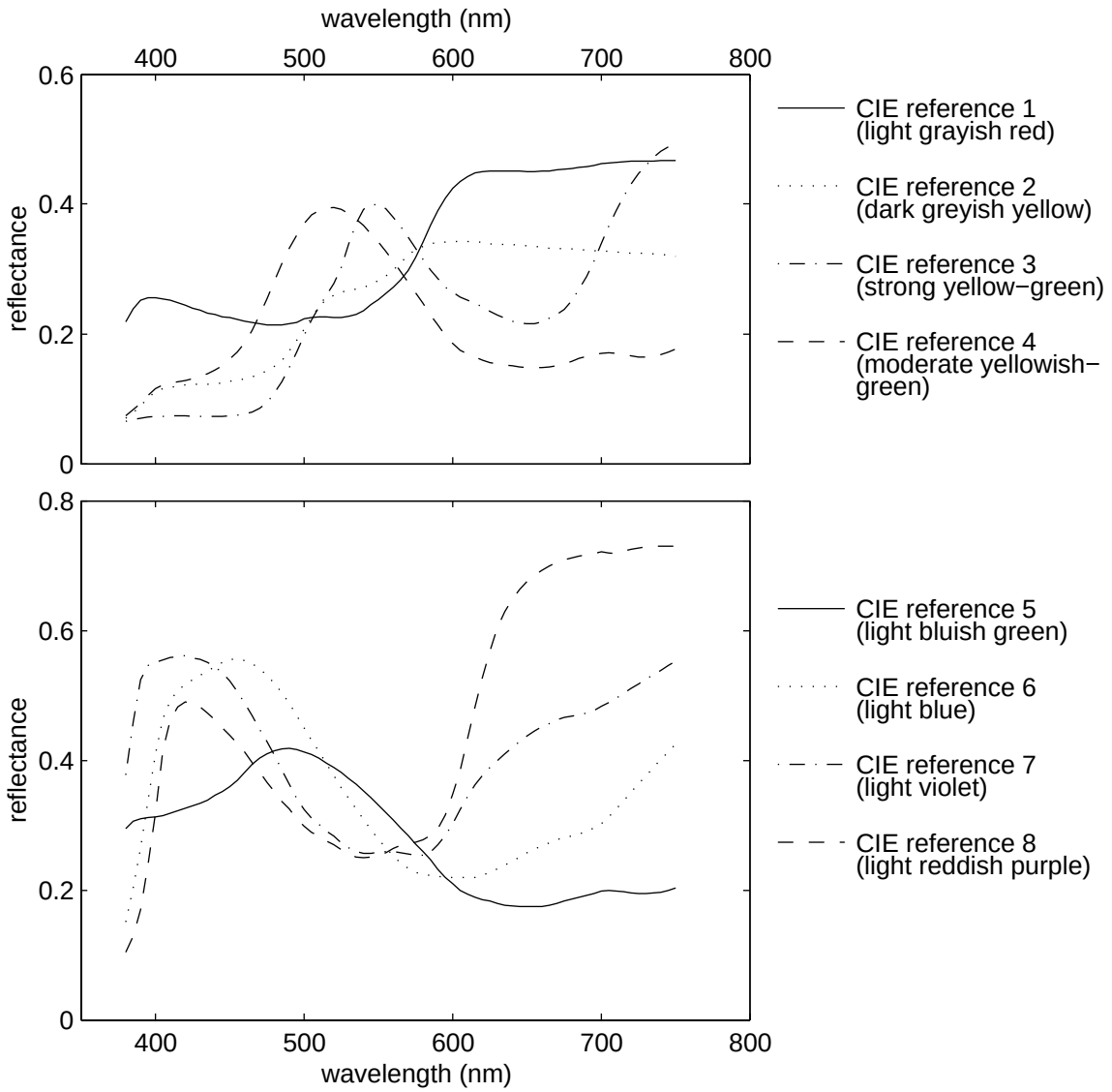

Figure 4. Reflectance of the CIE standard CRI test samples. These standard test colors used to determine the degree to which a test illuminant elicits color appearance in reflected objects similar to a reference illuminant.

2.2 Color Temperature and Correlated Color Temperature

Another metric frequently used to characterize white light sources is color temperature.

Color temperature is used to characterize spectra that are similar (in the visible range) to

Planckian radiators (Robertson, 1968). Planckian radiators are idealized[8], but many radiators in nature are nearly Planckian (e.g., the sun, tungsten filament lamps). The spectral radiancy of a Planckian radiator depends only on its temperature, and can be characterized by Planck's law (Halliday and Resnick, 1986):

$$R(\lambda,T) = \frac{2\pi c^2 h}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}$$

where:

$k = 1.381*10^{-23} J/K$ (Boltzmann constant)
$h = 6.626*10^{-34} J \cdot s$ (Planck constant)
$c = 2.998*10^{17} nm/s$ (speed of light in vacuum)
T = temperature (in Kelvin)
λ = wavelength (in nm)

Figure 5 is a plot of the chromaticity diagram showing the blackbody locus, the set of points in a chromaticity diagram that represent the chromaticities of spectra of Planckian radiators (blackbody radiators) at different temperatures.

For spectra that are not well approximated by a Planckian radiator (e.g., fluorescent lamps, some light emitting diode combinations) but with chromaticities near the blackbody locus, correlated color temperature (CCT) is used. First suggested by Judd (1936), CCT is the temperature of the Planckian radiator with chromaticity nearest the chromaticity coordinates of the spectrum on the CIE 1960 (*u,v*) diagram. Like CRI, CCT is a somewhat flawed metric (Borbely, Samson, and Schanda, 2001) that is nonetheless firmly established in the lighting industry.

---

[8] In the sense that their spectral radiancy depends only on temperature, and not, e.g., on material or surface properties.

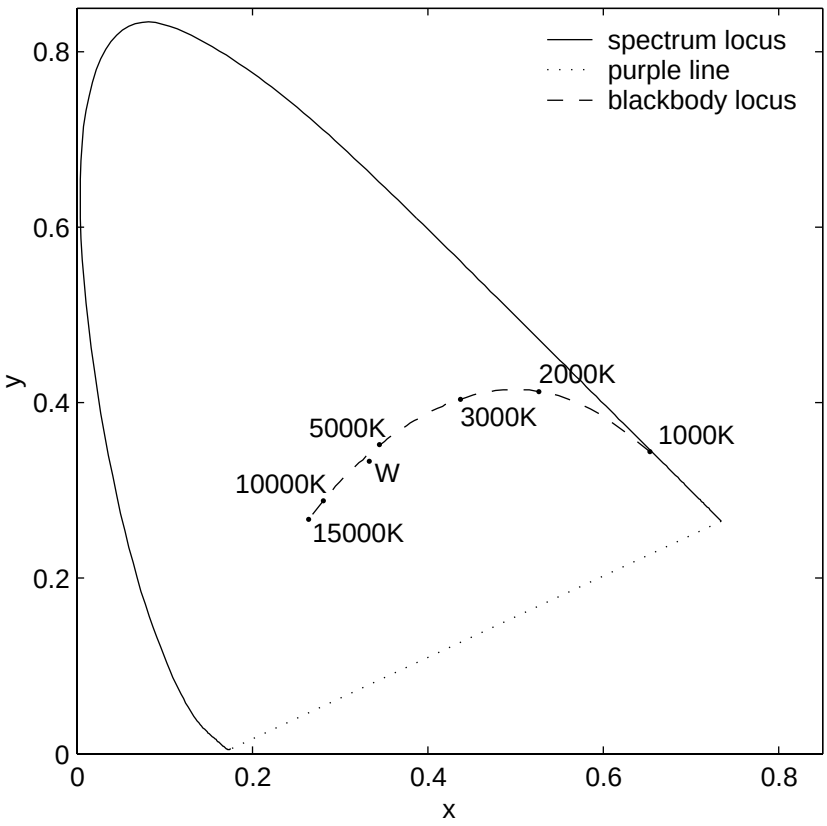

Figure 5. Chromaticity diagram with blackbody locus (the set of points in a chromaticity diagram that represent the chromaticities of spectra of Planckian radiators at different temperatures) and selected blackbody temperatures indicated.

3. Genetic Algorithms and Multiobjective Optimization

Classical optimization methods (either direct or derivative based) make strong assumptions about the search space (e.g., the fitness space is approximately quadratic; local minima or maxima are small or non-existent) that allow the optimal solution to be quickly determined. These methods are very powerful for the comparatively small set of problems for which their assumptions are known to be correct (or nearly so; Haupt, and Haupt, 1998). However, there are many problems where the character of the fitness space is unknown, or known to be unsuitable for classic optimization techniques. Moreover, there are many limitations to classical optimization techniques, including:

- the resulting solution may be highly dependant on initial conditions (i.e., identifying a local but not global optima)
- methods tailored to one class of problem are frequently unsuitable for other problems
- they may require an initial feasible solution, which can be difficult to generate
- integer (or discrete) variables are poorly handled
- problems must typically be described in a very specific way

Evolutionary algorithms (EAs) overcome these problems by employing techniques analogous to the ones that drive biological evolution; e.g., selection based on fitness, reproduction (sexual or asexual), and mutation (Michalewicz and Fogel, 2000). While there are some drawbacks to using EAs, they suffer from none of the limitations for classical optimization techniques mentioned above; hence, they are applicable to a broad variety of problems. The principal types of evolutionary algorithm include genetic algorithms (GAs), evolutionary programming, evolution strategies, and genetic

programming (Spears, et al., 1993; Koza 1995). Although all of these are driven by the principals of natural selection, and are therefore similar in overall operation, they differ substantially in the details of implementation.

This dissertation deals only with genetic algorithms. First proposed by Holland (1962; 1975), they did not really receive widespread attention from academia until the early nineteen eighties; the first conference dedicated to GAs was in 1985 (Grefenstette, 1985). Goldberg's (1989) book, which remains a standard text on the subject, substantially increased interest in the study of genetic algorithms.

The remainder of this chapter introduces the basic mechanics of genetic algorithms, introduces multiobjective optimization, and reviews previous work employing GAs for multiobjective optimization.

3.1 Genetic Algorithms

> *This preservation of favourable variations and the rejection of injurious variations, I call Natural Selection.*
> - Charles Darwin, 1896

Genetic algorithms are a general-purpose optimization method based on the theory of natural selection. GAs make no assumptions about the search space, so they can be applied to almost all optimization problems (Goldberg, 2002). However, GAs exchange applicability for speed – although they can be used on a wide variety of problems, they are typically slower to converge to a solution than algorithms designed for a specific problem.

Genetic algorithms employ a vocabulary borrowed from genetics (Michalewicz, 1996). A *chromosome* is an encoding of an individual solution to an optimization problem. A chromosome is composed of an ordered series[9] of *genes* (i.e., a specific gene always occupies the same position in a chromosome), with each gene corresponding to a specific characteristic of the problem[10]. A chromosome in its raw form (i.e., the encoded form that the GA manipulates) is known as the *genotype*; the solution that the chromosome represents (i.e., the solution decoded from the chromosome) is known as the *phenotype*. Each gene has a set of *alleles*, which are valid values for that gene. A *population* is composed of a collection (typically of fixed size) of chromosomes. An iteration of the algorithm during which a new population is produced is known as a *generation*.

The basic operation of a genetic algorithm is outlined in Figure 6. First, an initial population is generated (typically randomly). Then, the fitness of each chromosome is evaluated using a predefined fitness function. Fitness functions are used to evaluate the "goodness" of a chromosome, and can be either minimized or maximized, depending on the goal of the optimization. For example, consider a minimization problem represented by a chromosome with three genes, $[g_1, g_2, g_3]$, and the simple fitness function:

$$fitness = g_1^2 + 2g_2^2 - 4g_3$$

[9] Typically a vector, but sometimes arrays are used.

[10] Although there are some applications of diploid chromosomes (e.g., Green, 1994) the overwhelming majority of GA implementations employ haploid chromosomes (i.e., only one set of chromosomes per solution).

The fitness of the chromosome with allele values of [9, 3, 5] is 79 and the fitness of chromosome [2, -5, 2] is 46; therefore, [2, -5, 2] is a fitter ("more near optimal") solution for this minimization problem (although clearly not the optimal solution).

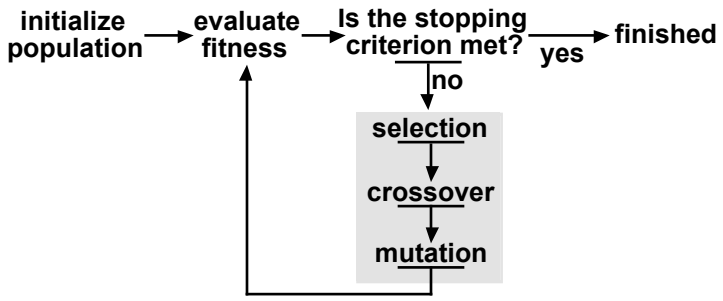

Figure 6. The GA optimization process. A population is initialized and the fitness of each chromosome is evaluated; if the stopping criterion is not met, selection, crossover, and mutation take place generating a new, presumably fitter, population, which is then returned to the fitness evaluation step; when the stopping criterion is met, the algorithm is terminated.

There are many possibilities for the stopping criterion, including:

- stopping after a fixed number of generations
- stopping after a predefined fitness has been achieved
- stopping when the rate of fitness improvement slows to a predefined level
- stopping when the population has converged to a single solution (e.g., when the maximum and mean population fitness is nearly the same)

A fixed number of generations and convergence are the most popular stopping criteria.

Selection chooses (frequently with replacement) chromosomes in the current population to be members of the following generation, either passed directly, or modified by crossover or mutation. The goal of selection is to favor relatively more fit chromosomes; those not selected (presumably, the least fit) die out. The three most commonly used selection schemes (Goldberg and Deb, 1991) are proportionate selection, rank selection,

and tournament selection. In proportionate selection ("roulette wheel" selection), the likelihood of selecting a chromosome is equal to the ratio of the fitness of the chromosome to the sum of the fitness of all chromosomes. One serious limitation of this method is that one comparatively very fit chromosome can very quickly overcome a population; rank and tournament selection are designed to overcome this problem. In rank selection, the population is sorted from best to worst fitness, and the probability of selection is some (linear or nonlinear) function of rank. In tournament selection, some small number of chromosomes (frequently two) are chosen at random, compared, and the fittest chromosome is selected; this process is repeated until sufficient chromosomes have been selected.

Crossover is the exchange of portions genetic material from a pair[11] of 'parent' chromosomes to produce a pair of 'children'. There are many ways to perform crossover (Michalewicz, 1996). The simplest method is single point crossover, where the chromosomes are split at a randomly selected point, and genes to the left of the split from one chromosome are exchanged with genes to the right of the split from the other chromosome, and vice versa. For example, consider these two parent chromosomes:

[*3, 4, 9, 1, 7, 6, 4, 8, 1*]
[**6, 3, 8, 1, 4, 3, 4, 6, 0**]

If the crossover point were between the third and fourth gene, these children would result:

[**6, 3, 8**, *1, 7, 6, 4, 8, 1*]
[*3, 4, 9*, **1, 4, 3, 4, 6, 0**]

[11] Typically, although crossover schemes with greater than two parents exist.

The specific methods of crossover employed for this work are described below.

Mutation (as it is typically implemented) is the assignment (with low probability of occurrence) of a random change to the allele value of one gene in a chromosome[12]. Mutation introduces new genetic material into the population and ensures that it is possible (over the course of the entire evolution) to search the entire solution space (Coello and Christiansen, 1999).

The newly generated population is passed to the fitness evaluation stage, and the cycle continues until the stopping criterion is met.

3.2 Multiobjective Optimization

Most real-world optimization problems have several, often conflicting, objectives. Therefore, the optimum for a multiobjective problem is typically not a single solution; rather, it is a set of solutions that trade off between objectives. This concept was first formulated by the Italian economist Vilfredo Pareto in 1896 (Tarascio, 1968), and it bears his name today. A solution is *Pareto optimal* if (for a maximization problem) no increase in any criterion can be made without a simultaneous decrease in any other criterion (Winston, 1994). The set of all Pareto optimal points is known as the *Pareto optimal front*.

[12] As with everything in GAs, there are no fixed rules: sometimes it is not a random change, and sometimes more than one gene is modified.

Figure 7 represents the objective space for an imaginary filtered lamp spectra, where one wants to maximize both luminous efficacy and CRI. The shaded area represents the feasible region (the region where solutions are possible). Solutions A and B are Pareto optimal: no increase in luminous efficacy can be made without an decrease in CRI, and vice versa. Solution C is *dominated* (not Pareto optimal): there are solutions with the same efficacy with a higher CRI (e.g., solution B), or with a higher efficacy for the same CRI (e.g., solution A). The heavy line indicates the Pareto optimal front – each point on it is non-dominated. Given the Pareto optimal front, an engineer can choose a solution based on other criteria (e.g., cost, ease of manufacture, parts availability).

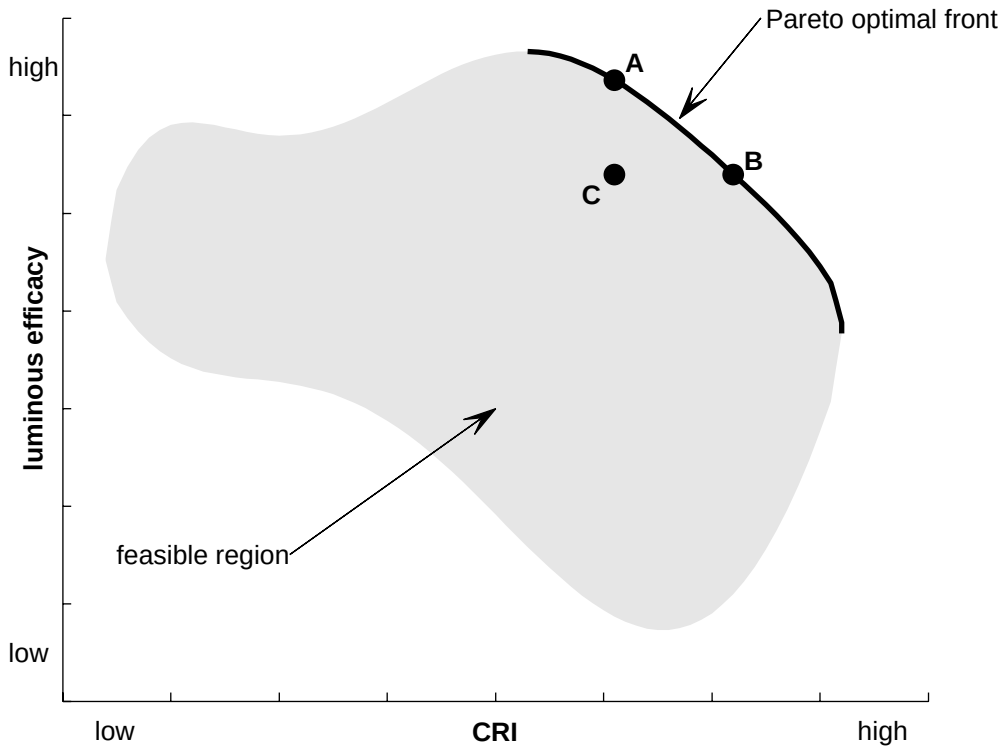

Figure 7. Concepts of multiobjective optimality. Solutions A and B are Pareto optimal: no increase in luminous efficacy can be made without an decrease in CRI, and vice versa. Solution C is dominated (not Pareto optimal): there are solutions with the same efficacy with a higher CRI (e.g., solution B), or with a higher efficacy for the same CRI (e.g., solution A). The heavy line indicates the Pareto optimal front - each point on it is non-dominated.

3.3 GAs for Multiobjective Optimization

The first real application of GAs for multiobjective optimization was by Schaffer in his 1984 Ph.D. dissertation (Deb, 2001; Schaffer, 1985); however it was not until the mid 1990s that the field really began to grow. The first international conference dedicated to multiobjective evolutionary algorithms was held in 2001 (Zitzler, et al., 2001).

Numerous approaches for using GAs for multiobjective optimization exist in the literature (Horn, Nafpliotis, and Goldberg, 1994; Fonseca and Fleming, 1995; Srinivas

and Deb, 1995; Zitzler, 1999). There are two common goals in all multiobjective GA implementations (although the details of implementation differ dramatically). First, to move the population toward the Pareto optimal front; and second, to maintain diversity (either in parameter space or objective space; Deb, 1998a) in the population so that multiple solutions can be developed. GA approaches to multiobjective optimization can be grouped into three categories (Coello, 1999; Zitzler, 1999; Fonseca and Fleming, 1995; van Veldhuizen, and Lamont, 1998; Coello, van Veldhuizen, and Lamont, 2002): approaches that use aggregating functions, non-Pareto based approaches, and Pareto-based approaches. Examples of each of these techniques are described below.

3.3.1 Approaches Employing Aggregating Functions

The simplest and most obvious approach to multiobjective optimization is to combine the objectives into one *aggregating function*, and treat the problem like a single objective optimization problem. Three major approaches that employ aggregating functions are the weighted sum approach, the constraint method, and goal programming.

The weighted sum approach combines k objectives using weights, w_i:

$$fitness = w_1 f_1(\overline{x}) + w_2 f_2(\overline{x}) + \ldots + w_k f_k(\overline{x})$$

(w_i are typically normalized such that $\sum_{i=1}^{k} w_i = 1$, although this is not necessary). Optimizing this function will result in one point on the Pareto-optimal front; therefore, multiple runs with different sets of weights are required to explore the whole of the Pareto-optimal front (Jones, et al., 1993). However, regardless of the weights employed, this technique will always miss concave portions of the Pareto-optimal front (Ritzel, Eheart, and Ranjithan, 1994; Liu, Begg, and Fishwick, 1998). Nonetheless, because of its

simplicity and computational efficiency, it is commonly used (e.g., Jakob, Gorges-Schleuter, and Blume, 1992; Wilson and Macleod, 1993)

The constraint method (e.g., Ranjithan, Eheart, and Liebman, 1992; Quagliarella, and Vicini, 1997) uses *k*-1 objectives as constraints, and the remaining objective, *h*, as the fitness function:

$$fitness = f_h(\overline{x})$$

subject to:

$$f_i(\overline{x}) \geq \varepsilon_i \qquad (1 \leq i \leq k, i \neq h);$$

the lower bounds, ε_i, are varied to find multiple points on the Pareto-optimal front. Unlike the weighted sum method, the constraint method is able to find points on any portion (including concave portions) of the Pareto-optimal front.

Goal programming, originally developed by Charnes, Cooper, and Ferguson (1955) for linear models, has been adapted for use with genetic algorithms for nonlinear problems (e.g., Deb, 1998b; Sandgren, 1994; Wienke, Lucasius, and Kateman, 1992; Wienke, et al., 1993). In its simplest form (Duckstein, 1984), fitness is calculated using:

$$fitness = \left|f_1(\overline{x}) - T_1\right| + \left|f_2(\overline{x}) - T_2\right| + \ldots + \left|f_k(\overline{x}) - T_k\right|$$

where T_i is a target value for the ith objective, which must be chosen by the researcher. A more general form (Haimes, Hall, and Freedman, 1975) is:

$$fitness = w_1(f_1(\overline{x}) - T_1)^p + w_2(f_2(\overline{x}) - T_2)^p + \ldots + w_k(f_k(\overline{x}) - T_k)^p$$

where p is typically 2, and w_i are weights; this is known as generalized goal programming (Ignizio, 1976, 1981) or target vector optimization (Coello, 1996). The principal strength of this method is that it is computationally efficient. Note that this method can result in

solutions that are not Pareto-optimal (Deb, 1998b; Duckstein, 1984) if the target values are within the feasible region[13]. Moreover, the researcher is required to specify weights (that frequently compare "apples and oranges") and targets for each objective, so considerable domain knowledge is frequently required (Deb, 1998b; Duckstein, 1984).

Multiobjective optimization using aggregating functions are generally computationally efficient; however, they only identify one point on the Pareto optimal front during each run. Hence, to explore the shape of the Pareto-optimal front, multiple runs are required. Moreover, as Zitzler (1999) states, "As the runs are performed independently from each other, synergies can usually not be exploited which, in turn, may cause high computation overhead." Thus, the practical utility of the computational efficiency of individual runs of techniques using aggregating functions is much reduced when the goal is to find the tradeoff between multiple objectives at multiple points.

3.3.2 Non-Pareto Based Approaches

Numerous GA approaches to multiobjective optimization that rely on neither aggregating functions nor the concept of Pareto-optimality[14] have been developed (Coello, 1999; Zitzler, 1999; Fonseca and Fleming, 1995; Coello, van Veldhuizen, and Lamont, 2002; Deb, 2001). Several of the major approaches are described here.

Schaffer (1984, 1985) developed the vector evaluated genetic algorithm (VEGA), based on Grefenstette's (1984) GENESIS program. In VEGA, each of k objectives is used, individually, to select a fraction of the total population (N) using proportionate selection.

---

[13] This is not a limitation *per se*; that is just the nature of goal programming.
[14] Approaches that employ the concept of Pareto-optimality are discussed in the following section.

These N/k subpopulations are mixed (shuffled), and crossover and mutation are applied. Like the weighted sum approach, this method can not find solutions on concave portions of the Pareto-optimal front (Richardson et al., 1989). However, because of its simplicity modified versions of VEGA have been used in a number of studies (e.g., Ritzel, Eheart, and Ranjithan, 1994; Cvetković, Parmee, and Webb, 1998; Tamaki et al., 1995; Tamaki, Kita, and Kobayashi, 1996), and it is particularly useful in the area of constraint handling[15] (Coello, 2000; Surry, Radcliffe, and Boyd, 1995).

Allenson (1992) and Lis and Eiben (1996) propose using gender (identified by a gene on each chromosome) to identify objectives. The objective function used for selection is based on the gender of the chromosome, and crossover requires one gender from each of k objectives to produce one child. Although this method can be used to successfully find Pareto-optimal solutions (Allenson, 1992; Lis and Eiben, 1996), it becomes computationally inefficient as the number of objectives increases because of the number of parents required to generate a child (Coello, 1999).

The weighted min-max approach was originally developed for linear models in the late 1960s (Coello, 1999), and was further developed by Osyczka (1978), Rao (1986), and Tseng and Lu (1990). Hajela and Lin (1992) were the first to apply this technique to genetic algorithms. They used weights for the objectives encoded in each chromosome combined with fitness sharing (after Goldberg and Richardson, 1987), mating restrictions, and a vector-evaluated approach similar to VEGA (Schaffer, 1984, 1985) to

[15] i.e., where the solution must satisfy several constraints

evolve a set of Pareto-optimal solutions during a single run. Other researchers (Coello and Christiansen, 1998; Coello, Christiansen, and Hernandez, 1998; Coello, Hernandez, and Farrera, 1997) have developed variations on Hajela and Lin's (1992) technique. This technique is computationally efficient and relatively simple to implement, although care has to be taken in the implementation to prevent premature convergence (Coello, 1996), and there are some parameters that typically have to be determined empirically.

3.3.3 Pareto Based Approaches

Goldberg (1989) proposed using the concept of Pareto dominance for ranking and selection combined with some kind of niching mechanism (to prevent convergence to one point on the Pareto-optimal front) to address the limitations of VEGA (Schaffer, 1984, 1985). This concept has been successfully implemented directly by several researchers (e.g., Hilliard et al., 1989; Liepins et al., 1990; Stanley and Mudge, 1995). Two commonly employed (Deb, 2001; Coello, van Veldhuizen, and Lamont, 2002) variations on this idea are described below.

Fonseca and Fleming (1993) proposed an approach (referred to below as FFGA) that ranks chromosomes by the number of other chromosomes by which they are dominated. All chromosomes with the same rank are given the same fitness (so that they are sampled at the same rate), which has the unintended effect of creating large selection pressure, which may lead to premature convergence (Goldberg, and Deb, 1991). To temper this tendency, fitness sharing is implemented based on objective function values. While this method is relatively simple to implement, it is quite sensitive to the sharing parameter and it is arguable (Deb, 1999) that sharing should be implemented in parameter space rather than objective space. The FFGA has been employed successfully by many

researchers (e.g., Chipperfield and Fleming, 1995; Aherne, Thacker, and Rockett, 1997; Tan, and Li, 1997).

Horn and Nafpliotis (1993; also Horn, Nafpliotis, and Goldberg, 1994) proposed the niched Pareto genetic algorithm (NPGA), which uses Pareto dominance to decide the outcome of tournament selection. Two chromosomes are selected at random, and compared to a randomly selected set of (typically 10) chromosomes. If only one of the original pair is not dominated, it is the winner; otherwise (i.e., in the case of a tie), the outcome is decided using fitness sharing (Goldberg and Richardson, 1987). While it is computationally efficient (because selection uses only a subset of the population to determine Pareto dominance), Coello (1999) notes that "its main weakness is that besides requiring a sharing factor, this approach also requires a good choice of the size of the tournament to perform well, complicating its use in practice." Nonetheless, the NPGA has been employed successfully by numerous researchers (e.g., Belegundu et al., 1994; Poloni and Pediroda, 1997).

4. Optimizing for Luminous Efficacy and Chromaticity Only

While each SPD has only one pair of chromaticity coordinates, a certain pair of chromaticity coordinates can be produced by an infinite number of SPDs. Moreover, a light source might be filtered any number of ways to achieve a particular color - what will vary (among other properties, such as color rendering) is the brightness of the light source and, therefore, the relative efficiency of the filtered light source. For any chromaticity coordinates achievable[16] by a light source, there is a way to filter the light source such that the efficiency at that color is maximized. Because the apparent color of a light is immediately obvious to the casual observer (and color rendering less obvious), it is of interest to be able to achieve a certain chromaticity at high efficiency, regardless of color rendering.

This chapter details a GA approach for determining how an arbitrary SPD may be filtered to achieve any set of chromaticity coordinates at maximum efficiency. This problem was undertaken as a stepping stone toward the full (chromaticity, efficacy and CRI) problem, presented in the following chapter. It is less complex, and the work of MacAdam (1935a) provides a convenient way to check how well the approach works.

The SPDs of three light sources employed in this chapter are plotted in Figure 8. Note that SPDs can be smooth and continuous (e.g., incandescent lamps, sulfur lamps), or spiky, with energy either spread throughout the visible spectrum (e.g., metal halide

[16] Not all light sources can achieve all chromaticities; a light source must have non-zero energy throughout the visible range to potentially achieve any chromaticity through filtering.

lamps, fluorescent lamps), or concentrated principally in one portion of the visible spectrum (e.g., high pressure sodium lamps, low pressure sodium lamps). Thus, one might expect very different filters for each lamp type for a given chromaticity coordinate.

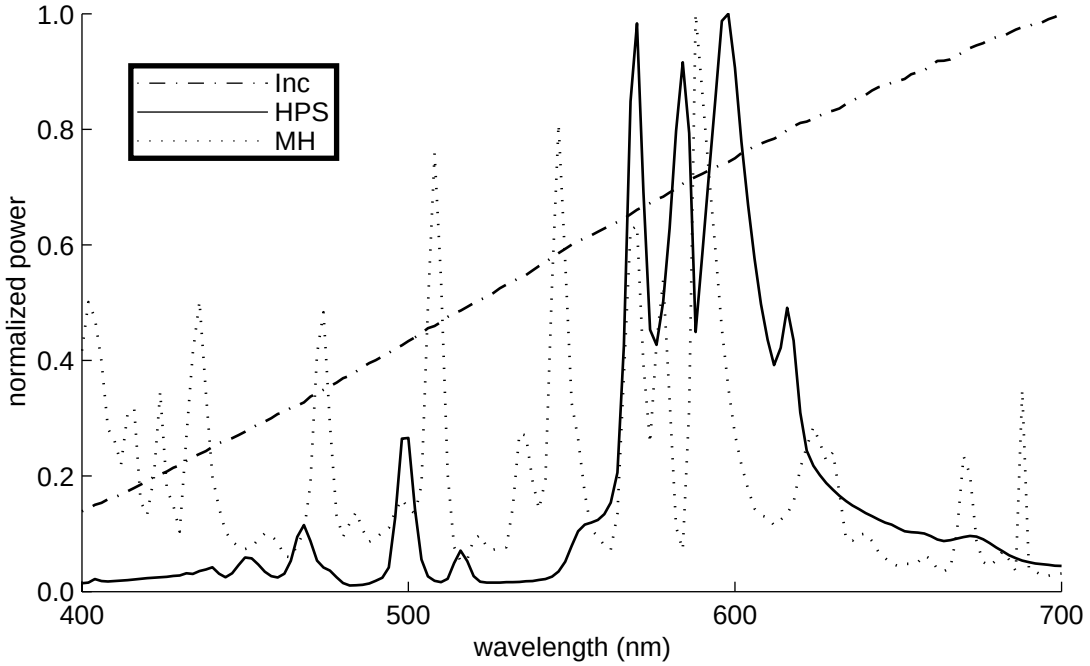

Figure 8. SPDs for Metal Halide, High Pressure Sodium and Incandescent lamps. Note that SPDs can be smooth and continuous (e.g., incandescent lamps, sulfur lamps), or spiky, with energy either spread throughout the visible spectrum (e.g., metal halide lamps, fluorescent lamps), or concentrated principally in one portion of the visible spectrum (e.g., high pressure sodium lamps, low pressure sodium lamps).

4.1 Genetic Algorithm Implementation

This optimization problem is unique, insofar as one is typically not interested in the whole of the chromaticity and efficacy Pareto optimal front (although that is developed below); rather, the maximum efficacy at a particular chromaticity (e.g., a point on the blackbody locus) is of interest. Thus, chromaticity will be "encouraged" to be near a

particular value[17] by using a penalty function: chromosomes near the target chromaticity will be penalized slightly or not at all, while chromosomes far from the target chromaticity will be strongly penalized. The fitness of any chromosome will be a combination of luminous efficacy and the chromaticity penalty (described in Section 4.1.2).

For this preliminary work, a modified version of the Genetic Algorithm Optimization Toolbox (GAOT; Houck, Joines, and Kay, 1995) was employed to operate the mechanics of the GA. Population sizes between 8 and 100 were experimented with; the results presented here are all for a population size of 50.

4.1.1 Chromosome Encoding

A floating-point representation was used for this problem. Michalewicz (1996) suggests that floating point representations tend to converge faster, reach more consistent results, and provide higher precision than binary representations. Other researchers give similar recommendations (e.g., Goldberg, 1990; Todd, 1997, Deb and Kumar, 1995)

The portion of the visible spectrum between 400 and 700 nm was partitioned in 151 bins, each 2 nm wide[18]. Each chromosome consisted of 151 genes, with each gene representing the transmittance of the filter in one of the bins. The order of the genes corresponded to

---

[17] The target-value approach used here us similar in spirit to the classic non-GA multiobjective optimization method known as goal programming (Cohon, 1978), which as also been adapted to GA applications (e.g., Wienke, Lucasius, and Kateman, 1992; Wienke, et al., 1993).

[18] The influence on chromaticity coordinates of the visible spectrum outside this range is trivial, less than 0.26% for a uniform SPD.

the order of the wavelength range in the spectrum; i.e., the first gene represented the 400 to 402 nm interval, the second gene represented the 402 to 404 nm interval, and so on. The 2 nm bin width was chosen as a compromise between smoothness and computational tractability. Valid allele values for each gene could range from zero to one.

4.1.2 Evaluating Fitness

The function,

$$fitness = \frac{Y_f}{Y_0} - 400 d_{cie}^2$$

where Y_f is the Y tristimulus value of the filtered SPD, Y_0 is the Y tristimulus value of the unfiltered SPD, and d_{cie} is the Euclidean distance from the target chromaticity coordinates (in the 1931 CIE system) to the chromaticity coordinates of the filtered SPD, was used to evaluate the fitness of solutions. The Y_f / Y_0 term is the relative efficiency of the filtered source (used throughout this document as a surrogate for luminous efficacy) that can range from 0 to 1. The d_{cie} term is scaled such that it was nearly zero near the target chromaticity (allowing the population some slack to search for more efficient solutions) and quite high away from the target chromaticity, penalizing chromosomes far from the target chromaticity. The scaling of the distance term was determined empirically.

This small area of low penalty around the color objective allows the efficiency to improve (in slight steps), even if color gets slightly poorer (also in small steps). The balance between the two objectives in this aggregating approach is important: if one of the objectives dominates the other, the population will become “stuck” at unreasonable

solutions. For example, Figure 9 is a plot of an solution found with a poorly scaled distance penalty; the color is right, but the efficiency is very low (.31), and the solution is likely too irregular to be economically (or even physically) manufactured. However, when the objectives are not poorly scaled, tradeoffs between color and efficiency can be made in the population, and much better solutions can be found. For example, Figure 10 is a plot of the transmittance of a filter with the same target chromaticity coordinates as for Figure 9; the efficiency is much higher (0.65), and the solution is (in principal) physically realizable.

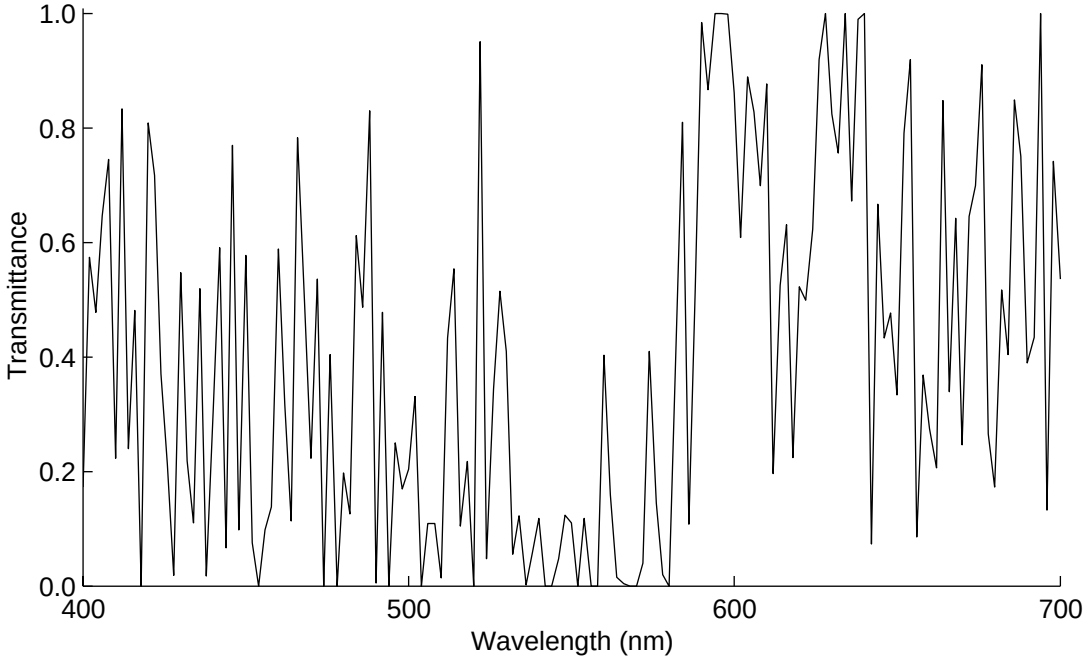

Figure 9. An poor solution for the incandescent lamp. Color is correct (for target), but efficiency is low (.31) and filter is physically unrealizable.

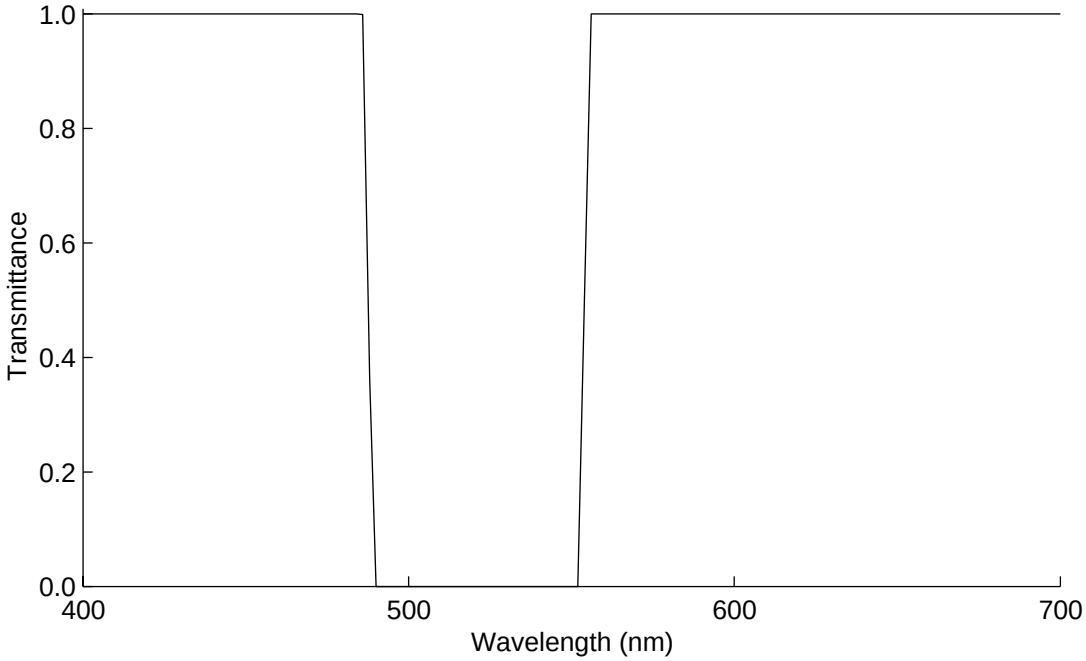

Figure 10. A good solution for the incandescent lamp (cf., Figure 9). Transmittance of a filter with the same target chromaticity coordinates as for Figure 9; the efficiency is much higher (0.65), and the solution is (in principal) physically realizable.

4.1.3 Selection

During each iteration 50 chromosomes are selected (with replacement) for crossover and mutation. Selection probability is rank based, with the probability of selecting the ith chromosome (when the chromosomes are sorted according to fitness), P_i according to:

$$P_i = \frac{q(1-q)^{r-1}}{1-(1-q)^n}$$

where:

q = probability of selecting the fittest individual
r = rank of chromosome (1 is best)
n = population size

Figure 11 is a plot of probability of selection against rank for the value of q used here (0.08).

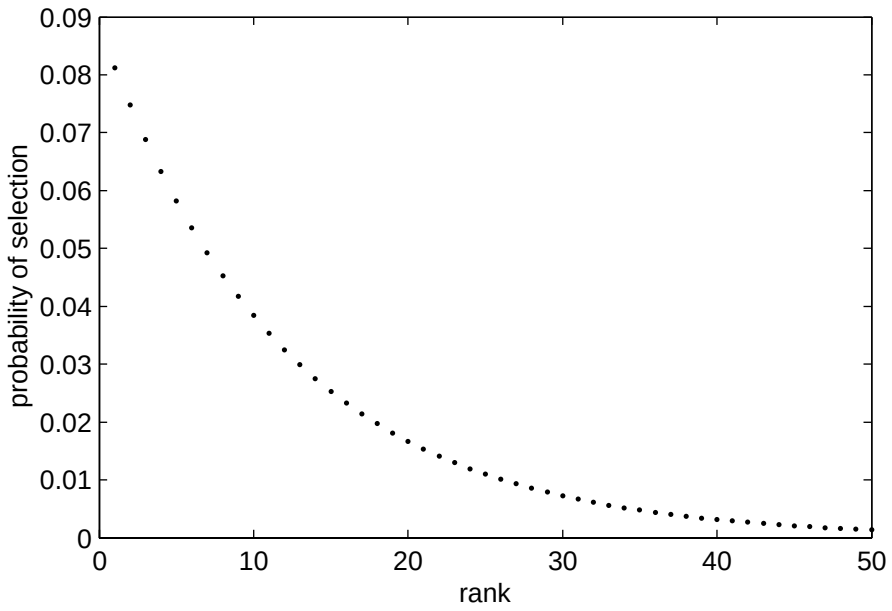

Figure 11. Probability of selection for each rank.

In addition to the regular selection process described above, the fittest chromosome in each generation is also passed on, unmodified by crossover and mutation (although also eligible for selection for crossover and mutation), to the next generation. This practice is known as "elitism", and is meant to preserve progress from generation to generation, and is well known to improve overall GA performance (Michalewicz, 1996; Mitchell, 1996).

4.1.4 Crossover

Three mechanisms of crossover were applied simultaneously: single point crossover (described above; 20 times per generation), arithmetic crossover (15 times per generation); and heuristic cross-over (15 times per generation).

Arithmetic crossover produces two complimentary linear combinations of the parents (Houck, Joines, and Kay, 1995):

$$\overline{X}' = r\overline{X} + (1-r)\overline{Y}$$

$$\overline{Y}' = (1-r)\overline{X} + r\overline{Y}$$

where:

$\overline{X}$ and $\overline{Y}$ are parent chromosomes
$\overline{X}'$ and $\overline{Y}'$ are child chromosomes
$r = U(0,1)$ [19]

Note that r is generated each time a pair of chromosomes is crossed, and that r is the same for all genes in both chromosomes crossed.

Heuristic crossover is based on interpolation, moving in the direction of the fitter chromosome (Houck, Joines, and Kay, 1995). If $\overline{X}$ is fitter than $\overline{Y}$ then (using the same notation),

$$\overline{X}' = \overline{X} + r(\overline{X} - \overline{Y})$$
$$\overline{Y}' = \overline{X}$$

If any gene in $\overline{X}'$ violates its allele limits, a new r is generated and a new child produced. This is repeated at most three times until either a new child is generated, or the parents are passed unmodified to the new population.

4.1.5 Mutation

Optimal chromosomes developed using this fitness function tend to have two properties in common. First, many gene values are exactly at the limits of the allele (i.e., 100% transmission or 0% transmission) in relatively fit chromosomes. Second, adjacent genes have nearly the same value (smoothness). Moreover, from a colorimetric perspective, the

[19] A uniformly distributed random variable between 0 and 1.

portion of the visible spectrum at either extreme has very little effect on efficiency; thus, there is little selection pressure applied by these regions (i.e., the central portion of the visible spectrum exerts much more influence on fitness than either end). This suggests several methods of mutation specific to the spectrum optimization problem that might be expected to produce a substantial decrease in the number of generations required to converge to a good solution.

For example, consider the filter in Figure 12. Because region A has very little influence on efficiency, there is not much pressure to either smooth out the genes, or move them toward a boundary. In region B, there is clearly a notch developing, but many generations may be required to smooth it out. Region C is near but not at the boundary, and, like region A, has little effect on fitness. The following three problem-specific mutation methods were designed to address these issues.

Boundary chunk mutation (BCM) selects a random contiguous portion of the chromosome (up to 10% of the total length) and sets it to the one of the allele limits (either 1 or 0). This mutation was expected to be effective because many of the genes of fit solutions were at the maximum or minimum value (i.e., either 100% transmission, or 0%). Moreover, because most of the genes for any color could be expected to be at 100% transmission, the selection of which boundary to mutate to was biased slightly: 65% of the time it went to 1, 35% of the time it went to 0. Figure 13 shows BCM applied to part of region A of the solution depicted in Figure 12.

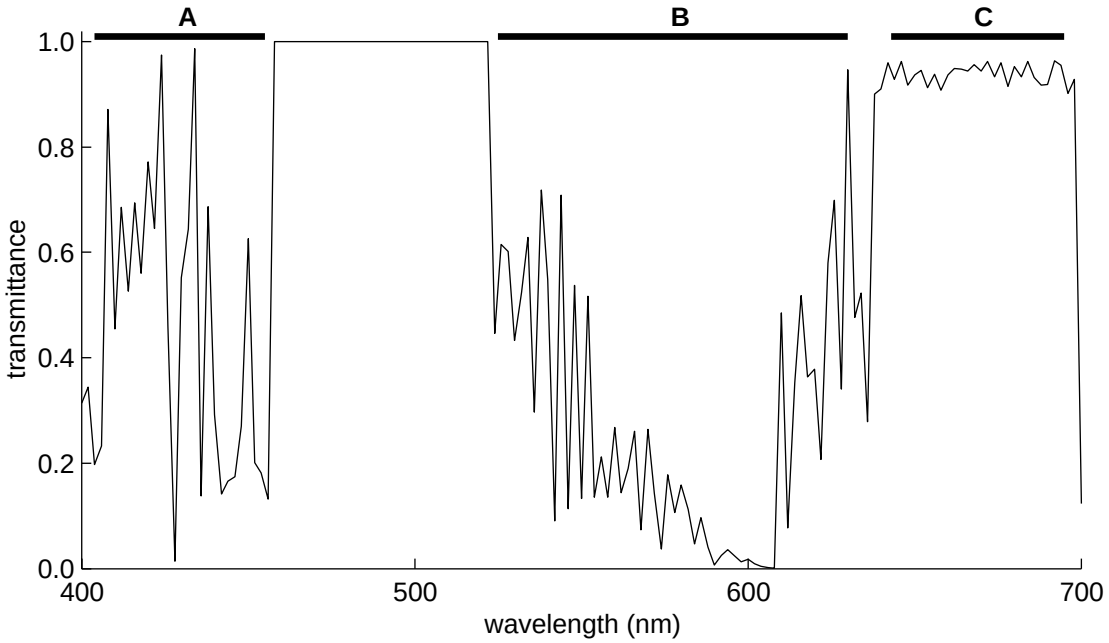

Figure 12. A sub-optimal filter. Because region A has very little influence on efficiency, there is not much pressure to either smooth out the genes, or move them toward a boundary. In region B, there is clearly a notch developing, but many generations may be required to smooth it out. Region C is near but not at the boundary, and, like region A, has little effect on fitness.

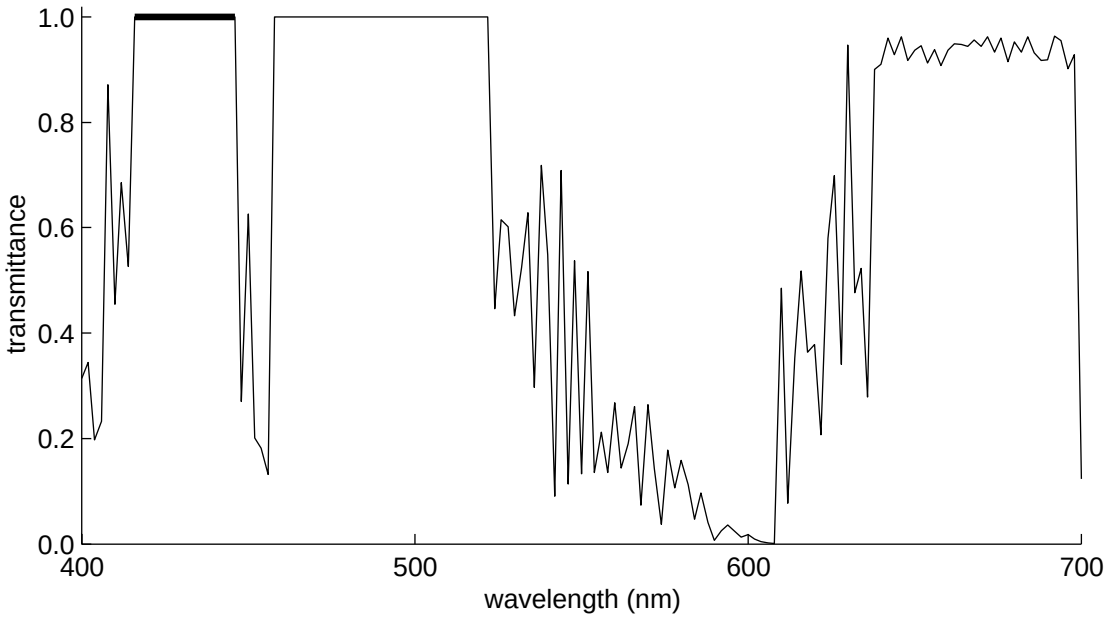

Figure 13. BCM applied to part of region A (bold line). BCM selects a random contiguous portion of the chromosome (up to 10% of the total length) and sets it to the one of the allele limits (either 1 or 0).

Push mutation (PM) selects a random contiguous portion of the chromosome (up to 20% of the total length) and scales the genes from their current value towards either 1 or 0 by a randomly chosen fixed amount (up to 20%). Figure 14 shows PM applied to part of region C of the solution depicted in Figure 12. Note that this might be effectively applied to region B as well, to help shape the developing notch.

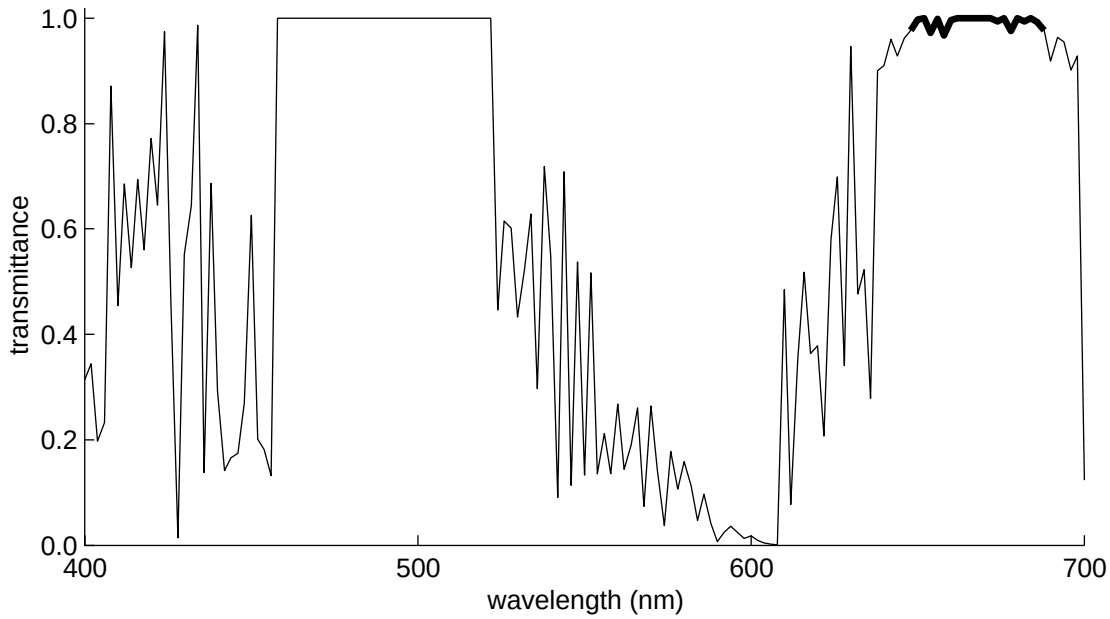

Figure 14. PM applied to part of region C (bold line). PM selects a random contiguous portion of the chromosome (up to 20% of the total length) and scales the genes from their current value towards either 1 or 0 by a randomly chosen fixed amount (up to 20%).

The smooth mutation (SM) selects a random contiguous portion of the chromosome (up to 20% of the total length) and smoothes it. Specifically, the value of each gene in the mutated portion is weighted by the value of its neighboring genes:

$$g_i = .2g_{i-1} + .6g_i + .2g_{i+1}$$

where g is a gene and i is the order of the gene in the chromosome. Figure 15 shows SM applied to part of region B of the solution depicted in Figure 12.

During each generation, the three types of problem specific mutation methods were applied five times each.

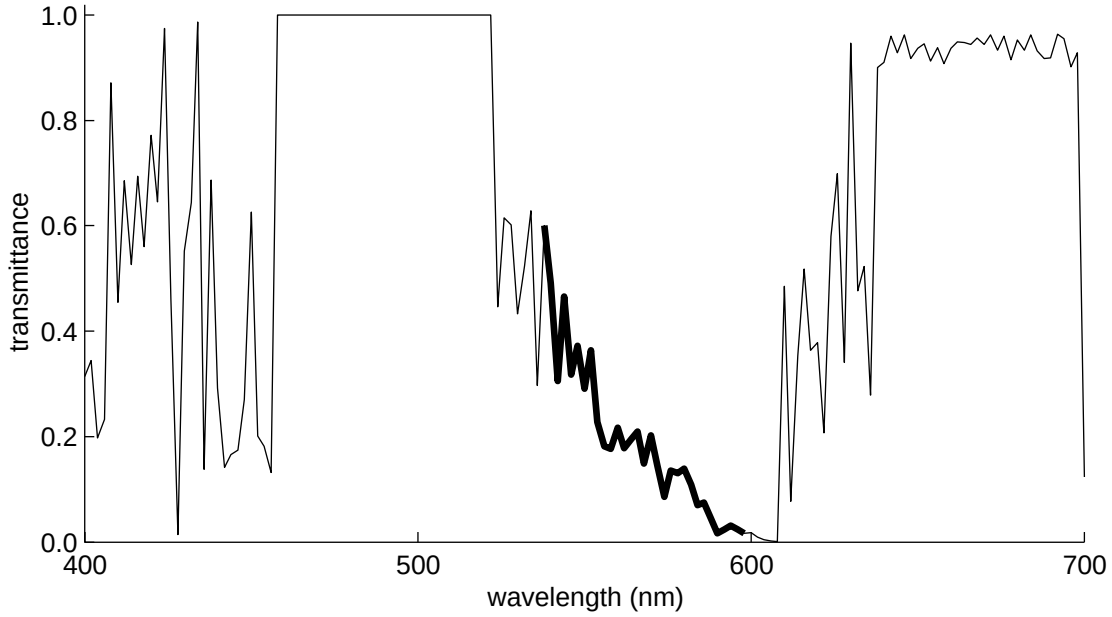

Figure 15. SM applied to part of region B (bold line). SM selects a random contiguous portion of the chromosome (up to 20% of the total length) and smoothes it by weighting it by the value of descent genes.

4.2 Efficacy & Chromaticity Results

Figure 16 presents the filters evolved by three typical runs of this technique (with stopping criterion of 1000 generations), one on each of the three spectra (MH, HPS, and incandescent), for the target chromaticity (x = 0.48; y = 0.32). This target chromaticity was selected to be about equidistant from the unfiltered spectra of the three lamps considered and achievable (albeit at low efficiency) by all three lamps. Also plotted are optimal filters produced for that color following MacAdam's (1935a) method.

The incandescent and metal halide filters are simple notches, very similar to the optimal notches found by MacAdam's (1935a) technique. The form of the HPS filter is quite different from the MacAdam optimal; however, the difference in efficiency between the two is less than 0.56%. This example illustrates two important features of genetic algorithms. First, the *exact* optimum is rarely found, although the difference is usually trivial (as it is here). Second, the GA approach is easily able to develop a variety of solutions[20] that have different form, but similar performance, allowing an engineer to choose from these (essentially) Pareto-optimal solutions based on criteria not encoded in the GA (e.g., ease of manufacture).

Table 2 gives the chromaticity coordinates, efficiency, and CRI of the filters plotted in Figure 16. Note that the CRI (which was not part of the fitness function) of the incandescent and MH spectra are reduced substantially by filtering.

Table 2. Chromaticity coordinates and efficiency of the filtered spectra.

	Unfiltered			Filtered			
Lamp	**x**	**y**	**CRI**	**x**	**y**	**CRI**	**Efficiency**
Incandescent	0.418	0.397	100	0.479	0.321	57	0.650
HPS	0.525	0.414	20	0.480	0.321	38	0.233
MH	0.369	0.382	66	0.479	0.321	0	0.510

[20] e.g., other runs for this problem have produced notches for the HPS lamp similar in form to the MacAdam optimal

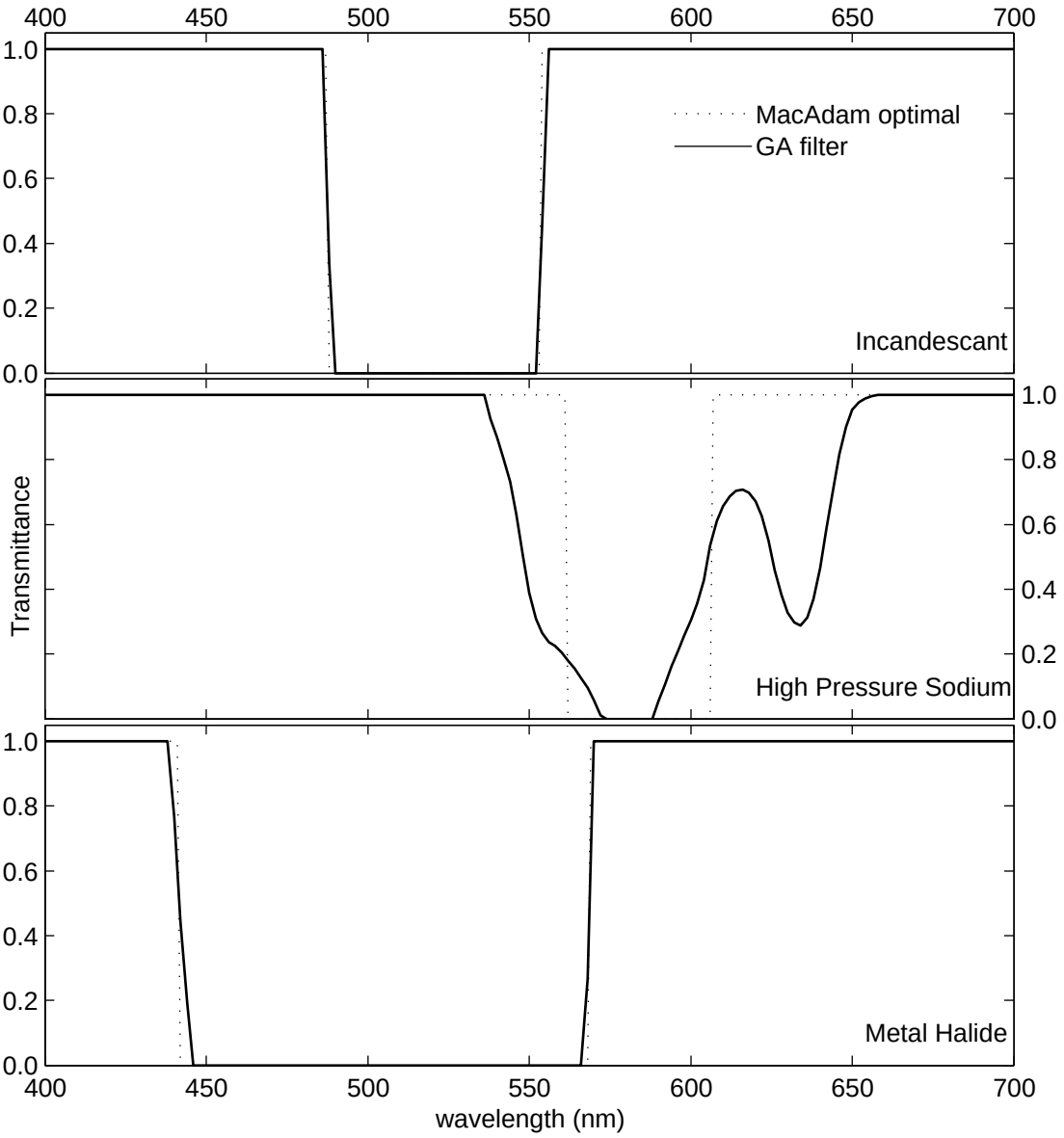

Figure 16. Filter transmittance for the three spectra (for the given chromaticity coordinates). The incandescent and metal halide filters are very similar the optimal notches found by MacAdam's technique; while the form of the HPS filter is quite different from the MacAdam optimal, the difference in efficiency between the two is less than 0.56%.

Figure 17 shows the progress of the fittest chromosomes through the chromaticity diagram for typical runs using each lamp spectra. The filled symbol shows the chromaticity coordinates of the unfiltered lamp. The line leading from each symbol is the chromaticity coordinate of the lamp filtered by the filter encoded in the best chromosome

each time a new best solution is generated. The inset figure in the lower left corner shows the CIE chromaticity coordinates plotted (not to the scale of the main figure) in the x-y plane. The box in the inset figure shows the area plotted in the main portion of the figure. About 260 filtered spectra are plotted for each lamp. Note that only approximately the first 20 are distinguishable outside the "blob" at (.48, .32).

Figure 18 is a way to visualize the path of the population through the color and efficiency space. This figure plots efficiency against x and y chromaticity coordinates (individually), with each spectrum in a different subplot. The symbols represent the unfiltered chromaticity coordinates. The filled symbols represent the x chromaticity coordinate; the empty symbols represent the y chromaticity coordinate. The line leading from each symbol is the efficiency of the lamp filtered by the filter encoded in the best chromosome each time a new best solution is generated, plotted against the respective x or y chromaticity coordinate.

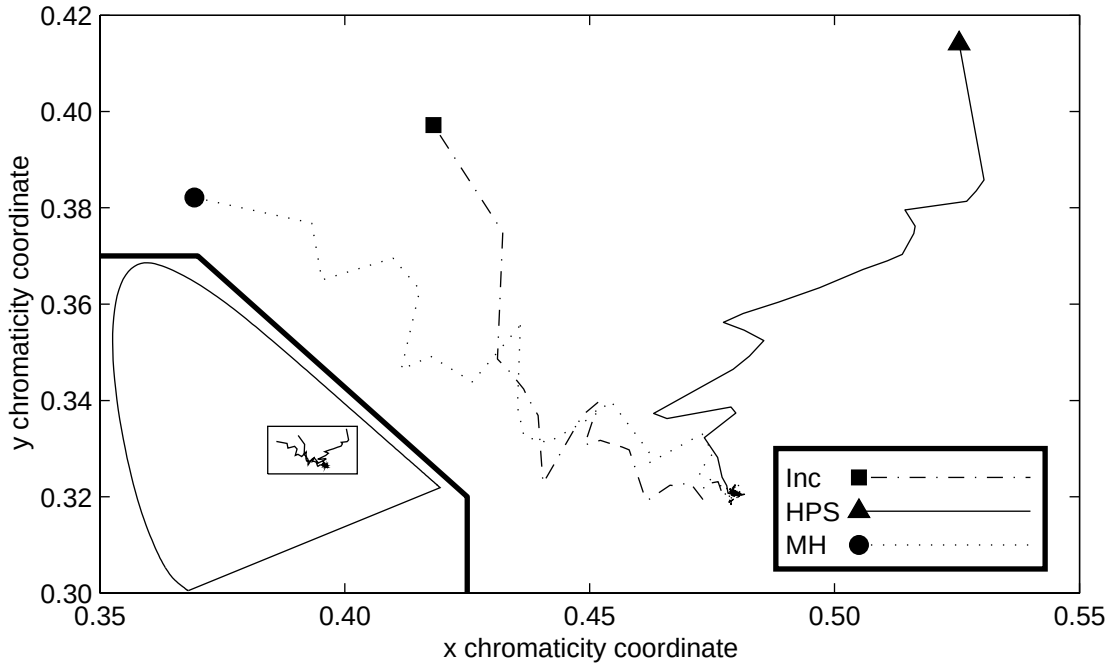

Figure 17. Path of the fittest chromosome through chromaticity diagram. The filled symbol shows the chromaticity coordinates of the unfiltered lamp. The line leading from each symbol is the chromaticity coordinate of the lamp filtered by the filter encoded in the best chromosome each time a new best solution is generated. About 260 filtered spectra are plotted for each lamp, but only about the first 20 are distinguishable outside the "blob" at (.48, .32).

Figures 17 and 18 plot the path of the fittest chromosome through chromaticity and efficiency space over time. Early on, relatively large steps toward the correct color are made, as shown in the region of Figure 17 where the individual points are distinguishable and the region of relatively horizontal movement in Figure 18 (following the first big step from the unfiltered SPD). After the neighborhood of the target color is found, efficiency is optimized in many small steps (the predominantly vertical, “squiggly” portion of Figure 18, at the opposite end of the line from the symbol). The small horizontal movements in this portion of the line represent tradeoffs between color and efficiency (i.e., a small deviation from the ideal color for a small gain in efficiency).

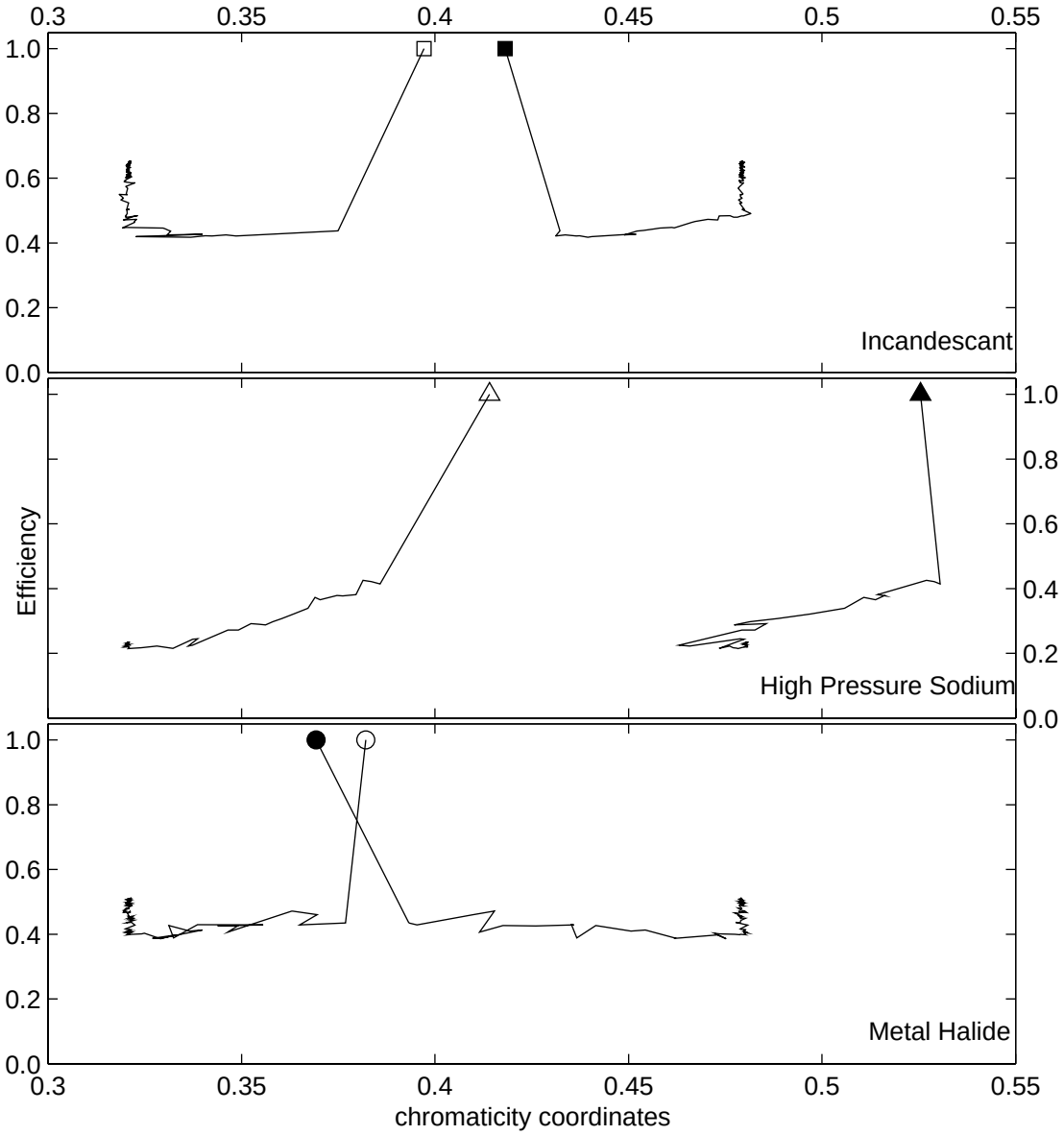

Figure 18. Path of the fittest chromosome through the search space. Efficiency is plotted against x and y chromaticity coordinates, with each spectrum in a different subplot; the symbols represent the unfiltered chromaticity coordinates (filled is x, empty is y); the line leading from each symbol is the efficiency of the lamp filtered by the filter encoded in the best chromosome each time a new best solution is generated, plotted against the respective x or y chromaticity coordinate.

The two portions of the path of the best chromosome through the chromaticity and efficiency space described above are a feature of the solution space (rather than a result of the way color and efficiency are scaled). It is easier to find the right color, because there are *many* different solutions to that portion of the problem. Meanwhile, it is quite

difficult to find the most efficient solution because there are very few (probably only one) "most efficient" solutions, and comparatively few solutions that are "very efficient". It is a tribute to the power of genetic algorithms that they are able to find an efficient solution at all.

4.2.1 The Pareto-Optimal Chromaticity-Efficiency Surface

The Pareto-optimal front for chromaticity and efficiency is a surface, because chromaticity has two dimensions (x and y), and efficiency provides the third. It is possible to determine the Pareto-Optimal chromaticity-efficiency surface using the method outlined in this chapter. This can be compared to the limits derived by MacAdam's (1935a) method, to determine how well the method outlined here performs.

A regularly spaced series of points (Figure 19) on the chromaticity diagram was cycled through as the target chromaticity coordinates for separate runs of the algorithm (with stopping criterion of 500 generations) in order to determine the nature of the tradeoff between efficiency and chromaticity.

Figures 20-22 are plots of the chromaticity-efficiency surface obtained using both methods[21]. The GA approach does a good job of defining the MacAdam limit of efficiency at any chromaticity, particularly where filtered efficiency is above 50% (which encompasses the area of likely interest for any industrial application). The GA performs less well at very low efficiencies; this is a relatively more difficult area to find the

[21] These data are plotted in the more perceptually uniform CIE 1976 (u',v') chromaticity space in Appendix A. The differences between MacAdam's limit and the results generated using the current method does not appear to be related to the nonuniformity of the 1931 CIE chromaticity space.

optimum filter in, so the stopping criterion (500 generations) was probably too restrictive for this region (i.e., if it had run longer, it would have performed better).

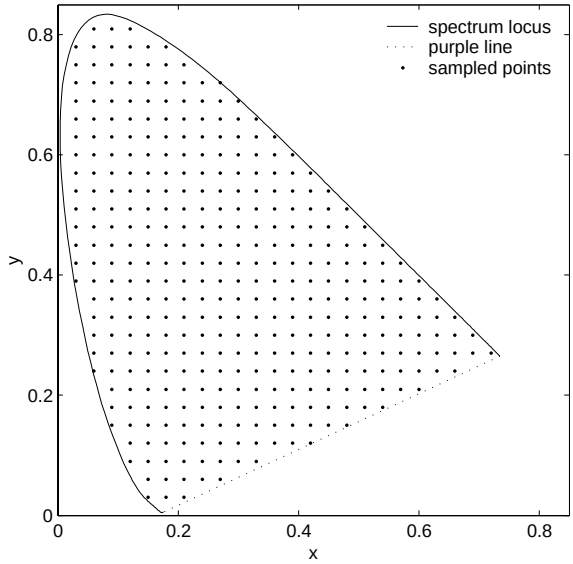

Figure 19. The points sampled for determining the chromaticity-efficiency surface.

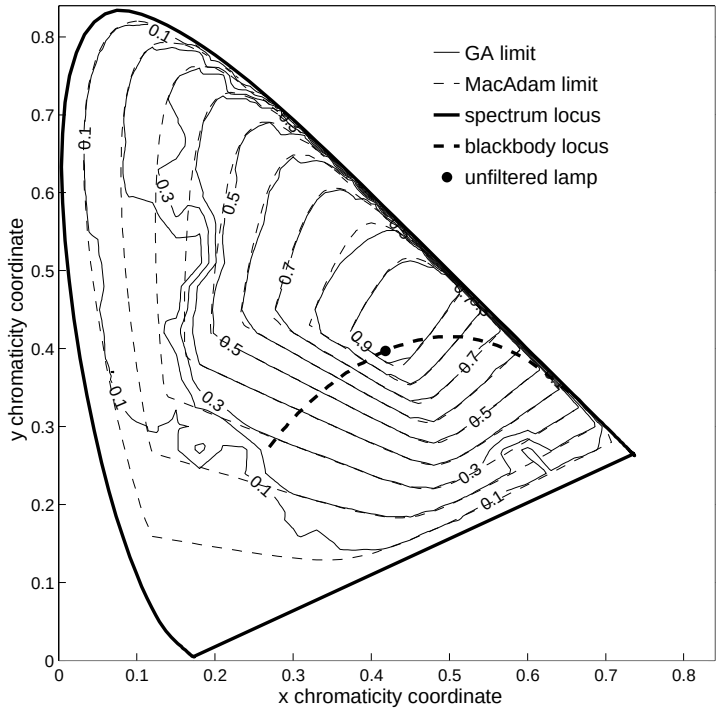

Figure 20. Efficiency contours for the incandescent source. Results from GA approach (solid line) are compared to the macadam limit (dashed line).

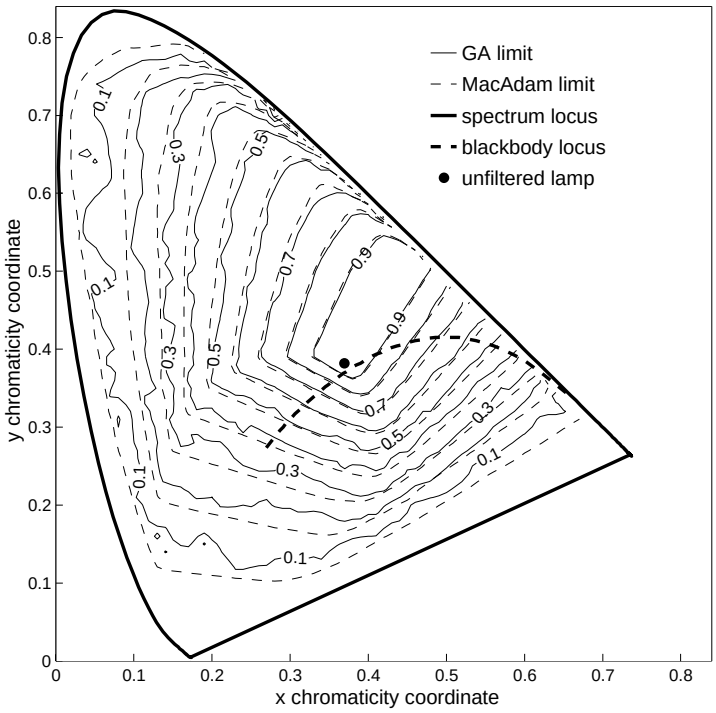

Figure 21. Efficiency contours for the MH source. Results from GA approach (solid line) are compared to the macadam limit (dashed line).

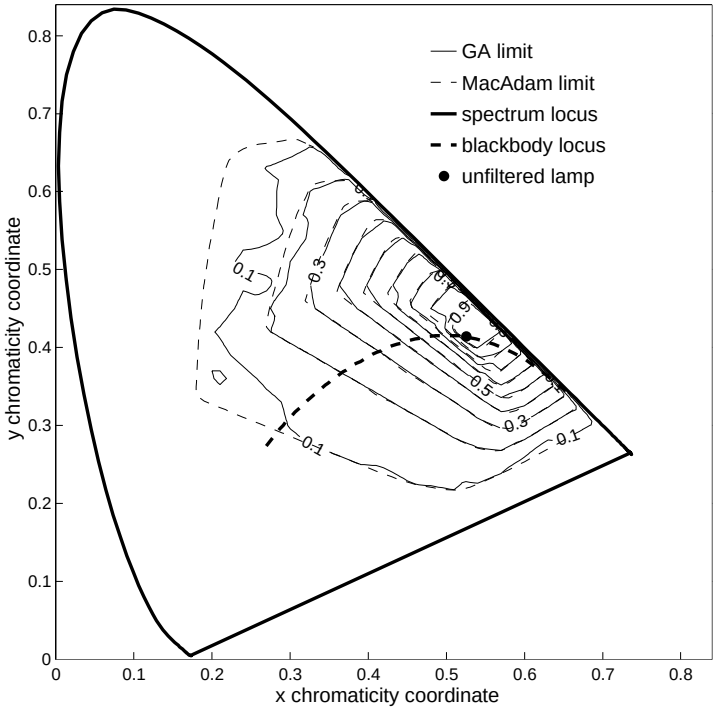

Figure 22. Efficiency contours for the HPS source. Results from GA approach (solid line) are compared to the macadam limit (dashed line).

The shape of the chromaticity-efficiency surface is consistent with what one might expect from the SPDs of the three sources. All three sources are capable of being filtered at a high (90%) efficiency so as to move the chromaticity coordinates of the filtered SPD in the "more saturated yellow" direction (toward ~580 nm on the chromaticity diagram). Meanwhile, there is very little slack to make the filtered SPD "more blue" at high efficiency. Figure 23 is a plot of "chromatic efficacy", the (normalized) amount of change in chromaticity when a small amount of energy is added at each wavelength (in turn) to a uniform (i.e., white) SPD, and the luminous efficacy at each wavelength (the Y tristimulus value). Note that wavelengths below 475 nm have a pronounced effect on chromaticity, but little effect on luminous efficacy: that is why moving in the yellow direction at high efficiency is possible, and moving in the blue direction is not.

The area of the chromaticity diagram achievable at any level of efficiency by the HPS source is much smaller than the area achievable by the other two sources, due to the HPS lamp having most of its energy in a relatively small region of the visible spectrum.

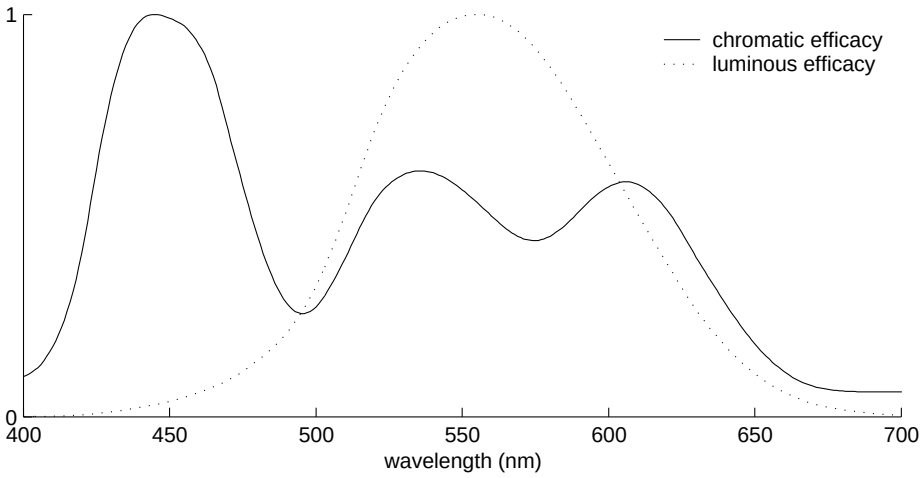

Figure 23. Normalized luminous efficacy and chromatic efficacy. Chromatic efficacy is the (normalized) amount of change in chromaticity for an incremental amount of energy added at each wavelength. Note that wavelengths below 475 nm have a pronounced effect on chromaticity, but very little effect on luminous efficacy.

4.3 Implications for Future Work

The results presented in this chapter are worthwhile insofar as they demonstrate two things. First, that the GA approach can produce results that are similar in fitness space, but quite different in solution space (e.g., Figure 16). Second, that this approach can develop solutions at or quite near known optima for this kind of problem (Figures 20-22), which suggests that GAs might be a reasonable approach for the three objective (chromaticity, efficacy, and CRI) case, for which there is no general solution.

However, this approach is quite inefficient, for two reasons. First, the chromosome encoding, while allowing very precise control (every 2 nm) over the location of changes in the filter transmittance, is quite long (for the same problem and level of precision, shorter chromosome encodings outperform longer representations; Michalewicz, 1996). Specialized mutation methods make up for this in part (e.g., by modifying blocks of genes), but a more compact representation is likely to be more efficient.

The second reason that this approach is inefficient is that one run only identifies a solution (or a couple of solutions) at a single chromaticity point. For applications like determining the Pareto-optimal front for chromaticity and efficiency, this approach is *extremely* inefficient, because many separate runs of the algorithm are required, and good building blocks of genetic material built up during one run are not shared in other runs (i.e., the GA is starting from scratch each run). Hence, an approach for developing solutions at many chromaticity points in parallel (and sharing good chunks of genetic material as these solutions are developed) is called for.

5. Main Results: Optimizing for Efficacy, Chromaticity, and CRI

> *The development of gas discharge lamps is primarily concerned with the improvement of (a) the total light output, and (b) the colour quality of light.*
>
> *This colour quality may be quantified using (i) the colour co-ordinates of the correlated colour temperature and (ii) the colour rendering indices.*
>
> - Koedam and Opstelten, 1971

There are two paths light from an electric light source may take to the eye: either directly from the source (apparent color), or reflected from an object (evaluated using CRI). The method outlined in the previous chapter addresses only the chromaticity of light directly from the source. While this is important (white[22] objects take on the hue of the illuminant), the degree to which an illuminant makes objects appear natural (e.g., compared to their appearance in daylight), that is, color rendering, is equally if not more important. As noted in the previous chapter (Table 2), the method described there (and MacAdam's method) do not consider color rendering, and frequently result in solutions that render colors poorly. Thus, if one is to consider the optimization of lamp spectra, it is necessary to account for luminous efficacy, chromaticity, and color rendering.

In the generic case of multiobjective optimization, one is interested in determining the whole of the Pareto-optimal front (e.g., the Pareto-optimal chromaticity-efficiency surface). However, in this specific case of spectrum optimization, the bulk of the Pareto-optimal front is known *a priori* to be uninteresting. In particular, chromaticities far from the Planckian locus are of little interest (because they appear non-white); moreover, even

[22] Characterized by high, uniform spectral reflectance over the visible range.

within chromaticities on the Planckian locus, there are some points that are "more interesting" (principally, from a marketing perspective) than others (e.g., at color temperatures that match older lamp technologies). Similarly, for the most part CRI values below 50 (if that) are of little interest (because the color rendering below CRI 50 is so poor), and in practice, CRI differences below 5 units are ignored (so a solution at 87 CRI and a solution at 83 CRI are, from a CRI perspective, essentially equivalent). Thus, a method for specifying the location of solutions in objective space (i.e., CRI, efficacy, and chromaticity) would be valuable.

5.1 Target Objectives Genetic Algorithm

Here, a novel approach to multiobjective optimization, the *target objectives genetic algorithm* (TOGA), is introduced. TOGA is a non-Pareto, non-aggregating function approach to multiobjective optimization that borrows concepts from goal programming and Schaffer's (1984, 1985) VEGA.

Given k objectives, TOGA requires the researcher to provide a set of c target vectors, $\overline{T}^c$, for k-1 objectives, where c is the number of optimal points the researcher wishes to find in one run of the GA. In the absence of domain knowledge to the contrary, taking all combinations of "interesting" points for each objective (considered separately) is the recommend method for developing T^c. In addition, k-1 scaling factors, w, are required (which can typically be determined empirically). During each generation, the objective values for each chromosome in the population are evaluated once[23], then the fitness

[23] Note that this property is particularly important in cases where the fitness function is very costly to evaluate; e.g., finite element analysis.

"from the perspective" of each target vector is calculated based on the objective values using (for a maximization problem):

$$F^c = f_1(\overline{x}) - w_1(f_2(\overline{x}) - T_1^c)^2 - w_2(f_2(\overline{x}) - T_1^c)^2 - \ldots - w_{(k-1)}(f_k(\overline{x}) - T_{(k-1)}^c)^2$$

where:

F^c is the fitness from the perspective of target vector c
$f_k(\overline{x})$ is the objective value of objective k
$w_{(k-1)}$ is a scaling factor for objective k
$T_{(k-1)}^c$ is the target value for objective k in target combination c

Given F^c, selection (with replacement) of a small, even number (e.g., 4, 6, or 8) of chromosomes is performed for each of c target combinations using a rank based procedure, with the probability of selecting the ith chromosome (when the chromosomes are sorted according to fitness), P_i according to:

$$P_i = \frac{q(1-q)^{r-1}}{1-(1-q)^n}$$

where:

q is the probability of selecting the fittest individual
r is the rank of chromosome (1 is best)
n is the population size

Crossover is performed separately on each of the c subpopulations formed. It is recommended that q be kept quite low (certainly less than .1), to promote exchange of genetic material from different locations on the optimal front (i.e., so that universally good building blocks developed in one region can spread rapidly to other regions). However, because of the low likelihood of preserving the best chromosome from each target combination, elitism is a critical feature of TOGA. During each generation the

chromosome with the maximum F^c for each of the c combinations[24] is passed on, unmodified, to the next generation. The method of mutation is not specified for TOGA (mutation should occur, but the particulars of implementation are not dictated by the approach).

The selection and elitism strategies are the key element in TOGA. By selecting from the perspective of each $\overline{T}^c$, subpopulations specializing in performance at that particular point on the optimal front are developed. However, diversity is maintained by having multiple $\overline{T}^c$, and by elitism. Moreover, by sharing good genetic building blocks developed on different regions of the optimal front, substantial efficiency is gained when compared to individual goal programming runs, which have to start from scratch.

TOGA has several advantages as an optimization method. First, TOGA is computationally very fast (i.e., the mechanics of the TOGA process do not require much computing power, particularly compared to Pareto based approaches). Second, TOGA can find points on a concave or convex Pareto front. Third, TOGA generates multiple optimal points during each run. However, TOGA also has a several limitations. First, TOGA is more efficient when the researcher has some domain knowledge and is able to select good combinations of objectives. Second, some trials are typically required to determine good scaling factors (although only k-1 scaling factors are needed). Finally, like goal programming, TOGA can results in points that are not on the Pareto-optimal

[24] Note that this might be the same chromosome for different target combinations, particularly during early generations.

front. This is not a limitation *per se*, rather it is the answer at that particular $\overline{T}^c$, but for practical purposes this answer may represent wasted computation[25].

5.2 TOGA Implementation

TOGA was implemented by defining target values for chromaticity and CRI, and leaving luminous efficacy unspecified. In particular, all combinations of five levels of CRI and seven levels of chromaticity on the blackbody locus (corresponding to levels of color temperature commonly encountered in the lighting industry worldwide) were combined (Table 3) to form the 35 target vectors, $\overline{T}^c$.

TOGA addresses both of the issues for multiobjective optimization outlined in Section 3.2: diversity is maintained by selecting for chromosomes near the $\overline{T}^c$, and selection pressure drives the population toward the Pareto-optimal front. Moreover, the selection method allows for crossover throughout the population (with greater likelihood of crossover between neighboring targets), so good "building blocks" of genetic material can spread rapidly through the population (which addresses one of the principal deficiencies of the method described in the previous chapter).

[25] Note that when multiple runs are performed, non-Pareto-optimal T^c can be removed, or repositioned.

Table 3. Target values for CRI and chromaticity, with corresponding color temperature.

Combination	CRI	CIE x	CIE y	Color Temperature
1	50	0.468	0.412	2600
2	50	0.437	0.404	3000
3	50	0.405	0.391	3500
4	50	0.380	0.377	4000
5	50	0.361	0.364	4500
6	50	0.345	0.352	5000
7	50	0.322	0.332	6000
8	60	0.468	0.412	2600
9	60	0.437	0.404	3000
10	60	0.405	0.391	3500
11	60	0.380	0.377	4000
12	60	0.361	0.364	4500
13	60	0.345	0.352	5000
14	60	0.322	0.332	6000
15	70	0.468	0.412	2600
16	70	0.437	0.404	3000
17	70	0.405	0.391	3500
18	70	0.380	0.377	4000
19	70	0.361	0.364	4500
20	70	0.345	0.352	5000
21	70	0.322	0.332	6000
22	80	0.468	0.412	2600
23	80	0.437	0.404	3000
24	80	0.405	0.391	3500
25	80	0.380	0.377	4000
26	80	0.361	0.364	4500
27	80	0.345	0.352	5000
28	80	0.322	0.332	6000
29	90	0.468	0.412	2600
30	90	0.437	0.404	3000
31	90	0.405	0.391	3500
32	90	0.380	0.377	4000
33	90	0.361	0.364	4500
34	90	0.345	0.352	5000
35	90	0.322	0.332	6000

Figure 24 shows the spectra employed in this chapter. Like the incandescent spectrum, the sulphur lamp spectrum is smooth and continuous, with energy throughout the visible portion of the spectrum. Figures 25 and 26 show the locations of the unfiltered spectra and target chromaticities on the 1931 CIE chromaticity diagram.

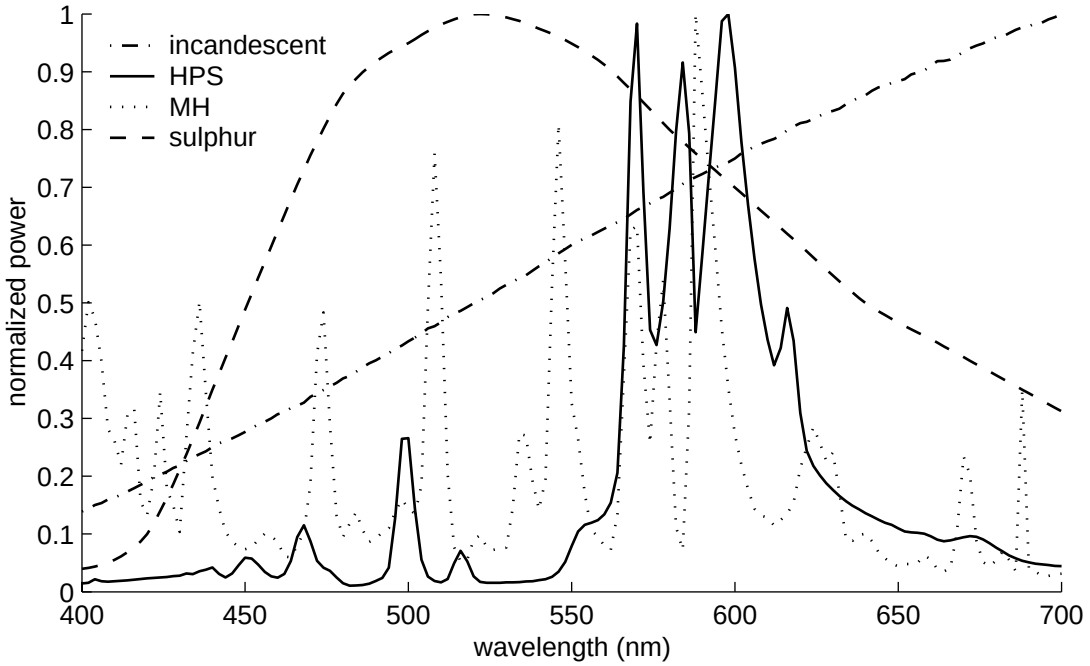

Figure 24. SPDs for incandescent, high pressure sodium, metal halide, and sulphur lamps. Note that like the incandescent SPD, the sulphur lamp SPD is smooth and continuous, with energy throughout the visible portion of the spectrum.

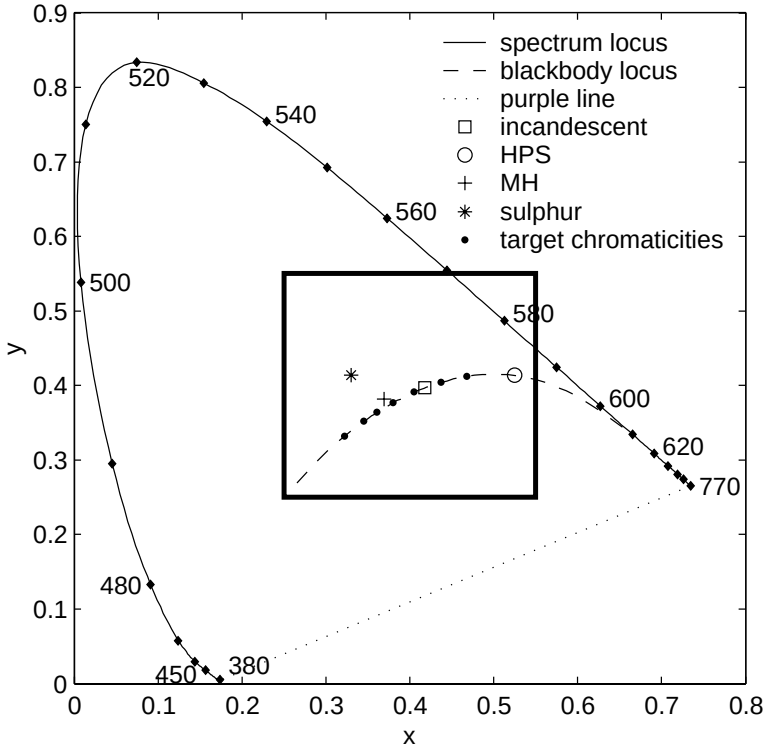

Figure 25. Positions of unfiltered spectra and target chromaticities on the CIE 1931 chromaticity diagram.

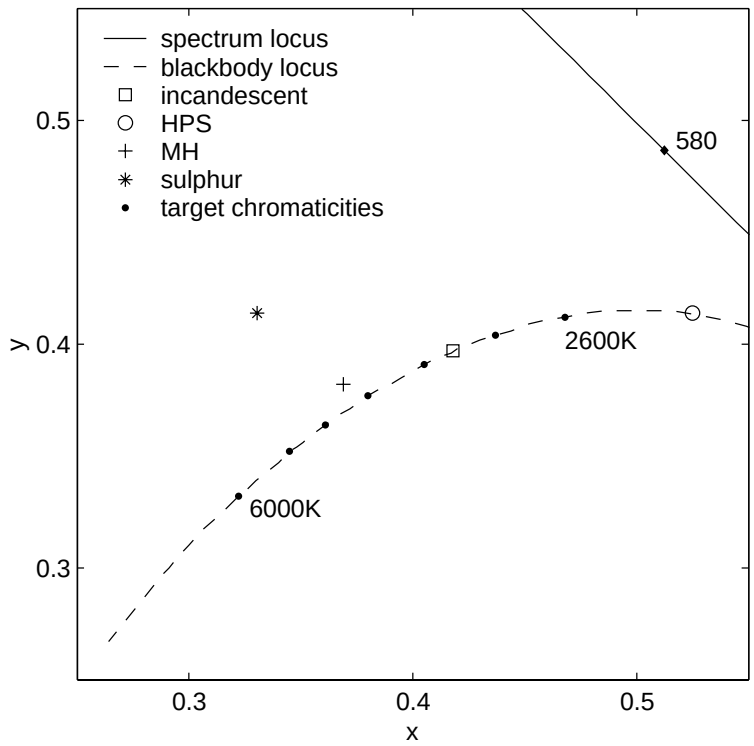

Figure 26. Boxed region of Figure 25.

Custom Matlab code was written to implement TOGA. The population size was 175 (five chromosomes allocated for each of 35 $\overline{T}^c$), and was randomly initialized. The algorithm was allowed to run for 500 generations before it was stopped.

5.2.1 Chromosome Encoding

A floating point representation was used for this problem, with allele values ranging from zero to one (although most were rescaled during decoding). Rather than encoding transmittance directly in the chromosome, a filter was constructed from four trapezoids (six and eight trapezoid encodings were examined, and found to produce filters similar in performance from four trapezoid encodings). While the use of trapezoids to construct the filter places some limitations on the shape of the filter, it is a good compromise between

computational tractability and flexibility: triangles would be easier to compute, but less flexible; a pentagon slightly more difficult to compute but more flexible; etc.

Each trapezoid was encoded by five parameters (Figure 27): base start position (a; scaled to 399 to 701 nm), left side width (b; scaled to 0 to 50 nm), top width (c; scaled to 0 to 50 nm), right side width (d; scaled to 0 to 50 nm), and height (e; a proportion, 0 to 1). Parameter 'a' determined where the leftmost edge of the trapezoid was; 1 nm on either side of the evaluated range (400 to 700 nm) was added to allow for trapezoids that did not affect the filter[26]. Parameters 'b' and 'd' were the width of the two sides; note that to produce right-angle notches, this width could be zero. Parameter 'c' was the width of the top of the trapezoid; a width of zero results in a triangle. Finally, parameter 'e' represented the height of the trapezoid; a height of zero meant no filtering at all, while a height of one meant 100% filtering (over the top of the trapezoid). The twenty parameters for all four trapezoids were concatenated to form the chromosome: $a_1 b_1 c_1 d_1 e_1 a_2 b_2 ... d_5 e_5$.

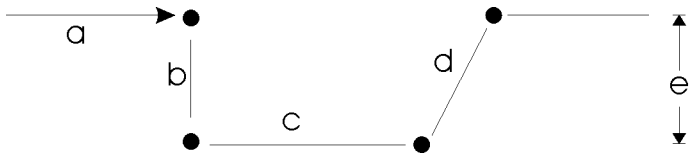

Figure 27. Trapezoid encoding. Parameter 'a' determined where the leftmost edge of the trapezoid was; parameters 'b' and 'd' were the width of the two sides (note that to produce notches, this width could be zero); parameter 'c' was the width of the top of the trapezoid; parameter 'e' represented the height of the trapezoid (note that a height of zero meant no filtering at all).

To decode the chromosome (converting it to a filter), the sum of the depths of each of the four trapezoids was evaluated at 3 nm intervals from 400 to 700 nm, and subtracted from

[26] This could also be accomplished by setting the height of the filter to zero.

a 100% transmittance filter (Figure 28). For example, the intermediate steps of decoding the chromosome:

[0.13 0.78 0.22 0.88 0.33 0.51 0.20 0.48 0.47 0.50 …
0.33 0.96 0.07 0.05 0.48 0.43 0.98 0.18 0.05 0.55]

is presented in Figure 28. The top four sets of axes show the individual trapezoids. The fifth axis shows the sum of the four trapezoids – note that where the sum is less than zero (dashed line), it is truncated. The bottom axis shows the final filter. Note that this method can produce anything from a fairly complex filter (Figure 28) to a simple notch.

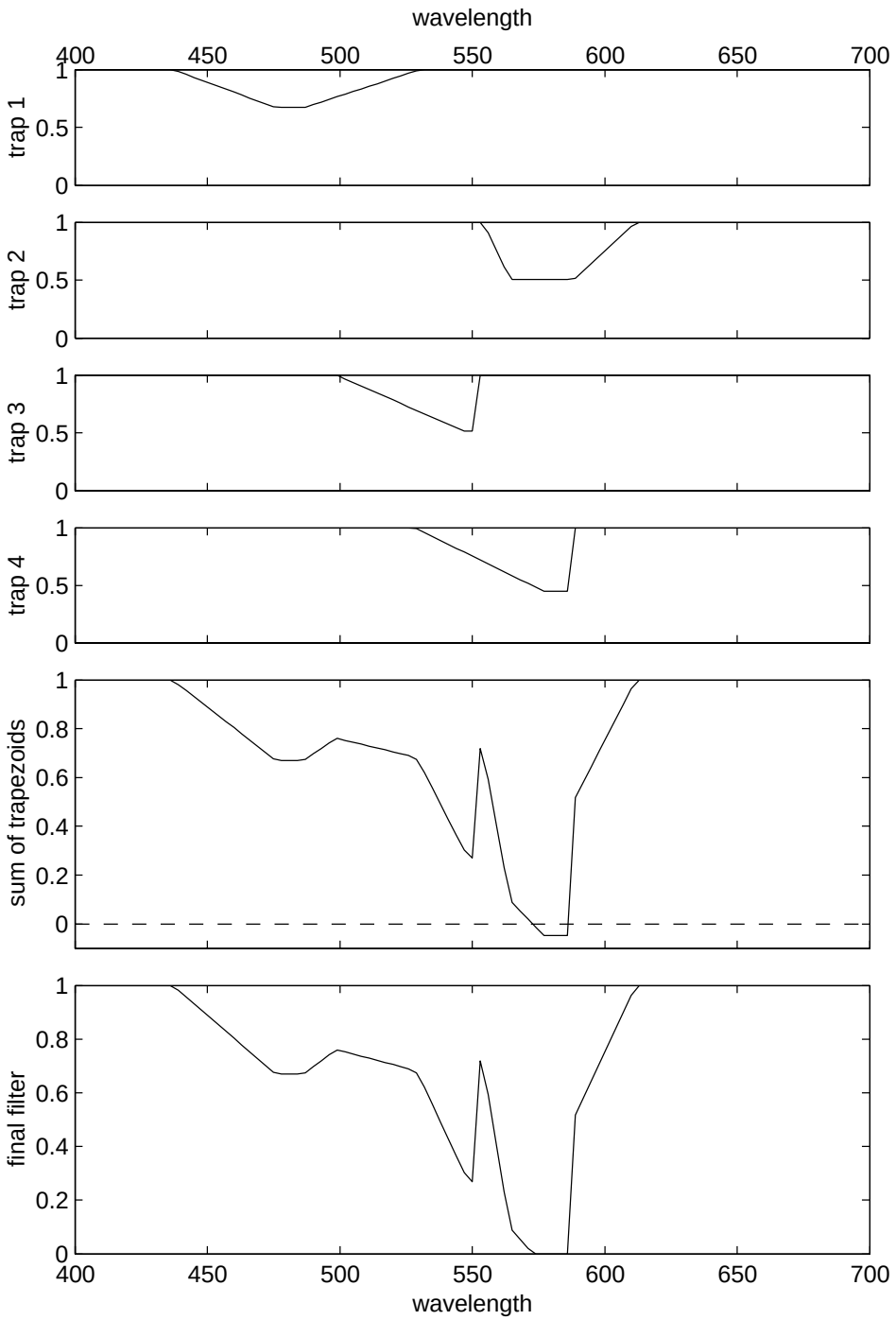

Figure 28. Intermediate steps to decoding a chromosome. The top four sets of axes show the individual trapezoids; the fifth axis shows the sum of the four trapezoids. Where the sum is less than zero (dashed line), it is truncated, resulting in the final filter shown in the bottom axis.

5.2.2 Evaluating Fitness

The function,

$$fitness = \frac{Y_f}{Y_0} - 900 d_{cie}^2 - \frac{d_{CRI}^2}{400}$$

where Y_f is the Y tristimulus value of the filtered SPD, Y_0 is the Y tristimulus value of the unfiltered SPD, d_{cie} is the Euclidean distance from the target chromaticity coordinates (in the 1931 CIE system) to the chromaticity coordinates of the filtered SPD, and d_{CRI} was the difference between the target CRI and the filtered CRI, was used to evaluate the fitness of solutions. The Y_f / Y_0 term is the relative efficiency of the filtered source (used here as a surrogate for luminous efficacy), which can range from 1 to 0. The d_{cie} and d_{CRI} terms are scaled such that they were nearly zero near the target (allowing the population some slack to search for more efficient solutions) and quite high away from the target, penalizing chromosomes far from the target chromaticity or CRI. The scaling of the distance terms was determined empirically.

5.2.3 Selection and Crossover

The fitness of each chromosome was calculated *from the perspective of each* $\overline{T}^c$. Then, for each $\overline{T}^c$, four chromosomes were selected (with replacement) for further crossover and mutation based on the rank-based method outlined in Section 4.1.3, with q set to 0.05. Chromosomes were crossed-over in the order they were selected (i.e., odd numbered chromosomes were crossed with even numbered chromosomes). Two mechanisms of crossover (single point and arithmetic) were applied, one every other generation.

Note that this selection and crossover scheme allows genetic material from very different places in fitness space to be combined (albeit at low likelihood). Crossover with chromosomes that are relatively near in fitness space is much more likely to occur. These two events in combination allow relatively fit building blocks to propagate easily throughout the population.

Elitism was also implemented, selecting the fittest chromosome from each $\overline{T}^c$ (for a total of five chromosomes selected for each $\overline{T}^c$).

5.2.4 Mutation

Five different mutation methods were used; two general purpose, and three problem-specific. The two general purpose (in the sense that these mutation methods were intended to promote diversity in the population, rather than address some specific feature of the problem) methods were "uniform mutation" and "normal mutation". Uniform mutation sets the value of a randomly selected gene to a uniform random value, and was applied five times per generation. Normal mutation sets the value of a randomly selected gene to a normally distributed random value, and was applied 50 times per generation. The mean for normal mutation was zero and the standard deviation was:

$$\sigma = 0.2(1 - \tanh(\frac{2.4g}{g_{\max}}))$$

where g is the current generation, and g_{max} is the maximum number of generations (Figure 29).

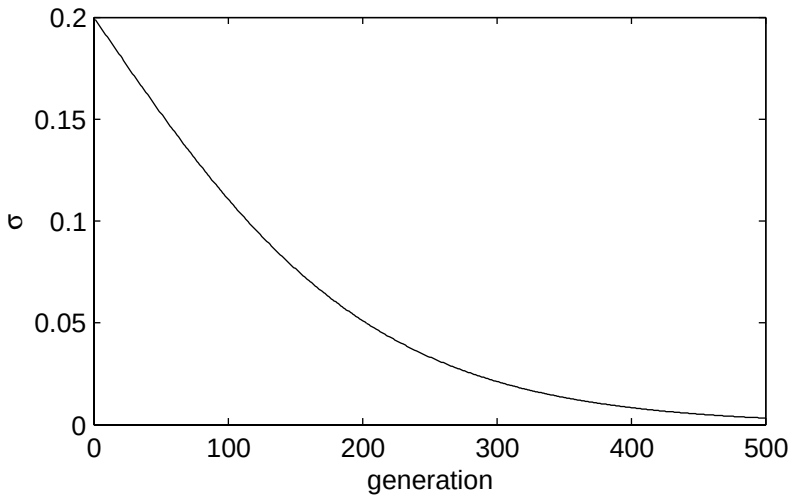

Figure 29. Standard deviation for normal mutation as a function of generation.

The three problem specific mutation methods modified particular features of the problem (based on speculation on what might lead to a good solution), and were intended to insert into the population genetic material that would lead to fitter chromosomes. Boundary mutation set a randomly chosen gene to the allele limit (either 0 or 1), and was applied to genes representing the 'e' parameter (depth) of the trapezoid, 15 times per generation. This mutation method served two purposes, either removing entirely the effect of a trapezoid (when set to zero), or helped convert the trapezoid to a full depth notch (when set to one).

Max mutation set a randomly chosen gene to the maximum allele value, and was applied to genes representing the 'a' parameter (start wavelength) of the trapezoid, 15 times per generation. This mutation method served to remove entirely the effect of a trapezoid, by pushing it beyond the evaluated region.

Zero mutation set a randomly chosen gene to the minimum allele value, and was applied to genes representing the 'b', 'c', and 'd' parameters, which correspond to first side width, top width, and second side width, respectively. In the case of parameter 'c', this mutation converts the trapezoid to a triangle. When affecting parameters 'b' or 'd', this mutation forms one side of the trapezoid into a right-angle notch. This mutation method was applied 15 times per generation per parameter.

5.3 Efficacy, Chromaticity & CRI Results

The main results are presented in two sections, one dealing with GA performance, and the other dealing with lighting. Results presented are for 150 separate runs for each of the four lamp spectra.

5.3.1 GA Results

For the method described here to be a good approach for determining the tradeoff between chromaticity, efficiency, and CRI for an arbitrary SPD, it is necessary that:

- chromaticity error (relative to the $\overline{T}^c$) be within acceptable limits
- CRI error (relative to the $\overline{T}^c$) be within acceptable limits
- the same (presumably global) optimum be found on most runs

These issues are assessed in this section.

Figure 30 is a plot of the distribution of chromaticity coordinate error ($[x,y]_{filtered} - [x,y]_{target}$) from 150 runs of the GA, for each $\overline{T}^c$ for the incandescent spectrum (results from the other spectra are presented in Appendix A). Each point represents error from an individual run. The horizontal and vertical bell curves are normal probability density functions with height normalized and scaled to fit within the

axes, and with the same mean and variance as the x and y chromaticity error, respectively. The dashed line represents a three standard deviation MacAdam ellipse (MacAdam 1942; MacAdam, 1943; Brown, et al., 1956) for the target chromaticity. A one standard deviation MacAdam ellipse is the boundary for just noticeable difference (JND) in chromaticity under ideal viewing conditions (adjacent fields, unlimited observation time, photopic conditions, foveal viewing); the JND will be larger for less than ideal viewing conditions (Brown, 1952, 1957; Brown, and MacAdam, 1949; Narendran, et al., 2000;). A four standard deviation MacAdam ellipse is the ANSI standard tolerance for lamp color difference (ANSI, 1996), although major manufactures stay within a three standard deviation limit (OSI, 2002). The chromaticity coordinates of solutions developed using this method are quite close to the target chromaticity, almost all falling well within industry standards of color tolerance (Table 4).

Table 4. Percentage of filtered spectrum chromaticity coordinates falling within MacAdam ellipses of different size.

Spectrum	MacAdam Ellipse σ			
	1	2	3	4
Sulphur	32	95	100	100
Incandescent	73	98	100	100
High Pressure Sodium	76	95	98	99
Metal Halide	22	75	96	99

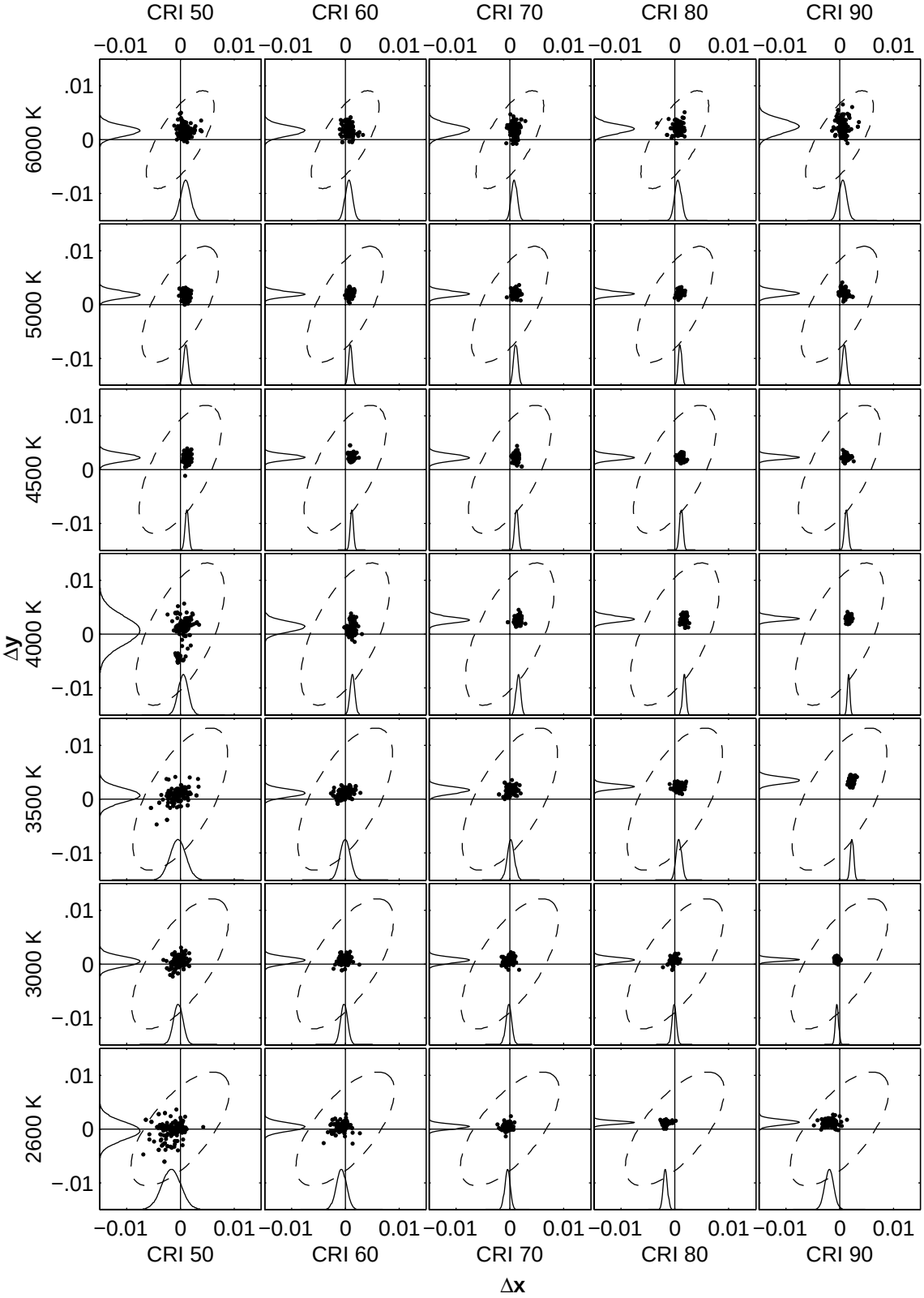

Figure 30. Distribution of chromaticity coordinate error by color temperature and CRI, incandescent spectrum.

Figure 31 is a set of boxplots of CRI error ($CRI_{filtered} - CRI_{target}$) for each $\overline{T}^c$. The box runs from the lower quartile (25th percentile) to the upper quartile (75th percentile), with a vertical line at the median. The "notch" at the median is a robust estimate of uncertainty in the median (analogous to the standard error of the mean). The whiskers (the lines extending from each end of the box) extend to 1.5 times the interquartile range (75th percentile minus the 25th percentile). If there is no data outside the whiskers, a dot is placed at the leftmost whisker; values beyond the ends of the whiskers (outliers) are plotted with a "+" symbol. The vertical dashed line represents zero error.

Five units of CRI is just perceptibly different (IESNA 2000). The CRI error for the filters developed here are quite small (well with industry limits) for the overwhelming source and $\overline{T}^c$ combinations, with the exception of the MH spectrum at color temperatures below 3500K. Overall, the CRI of solutions developed using this method are quite close to the target CRI, almost all falling well within industry standards of color tolerance (Table 5).

Table 5. Percentage of filtered spectra with CRI within one to five units of the target CRI.

	Absolute CRI Error				
Spectrum	**1**	**2**	**3**	**4**	**5**
Sulphur	70	97	100	100	100
Incandescent	82	97	99	100	100
High Pressure Sodium	85	97	99	100	100
Metal Halide	52	81	91	96	98

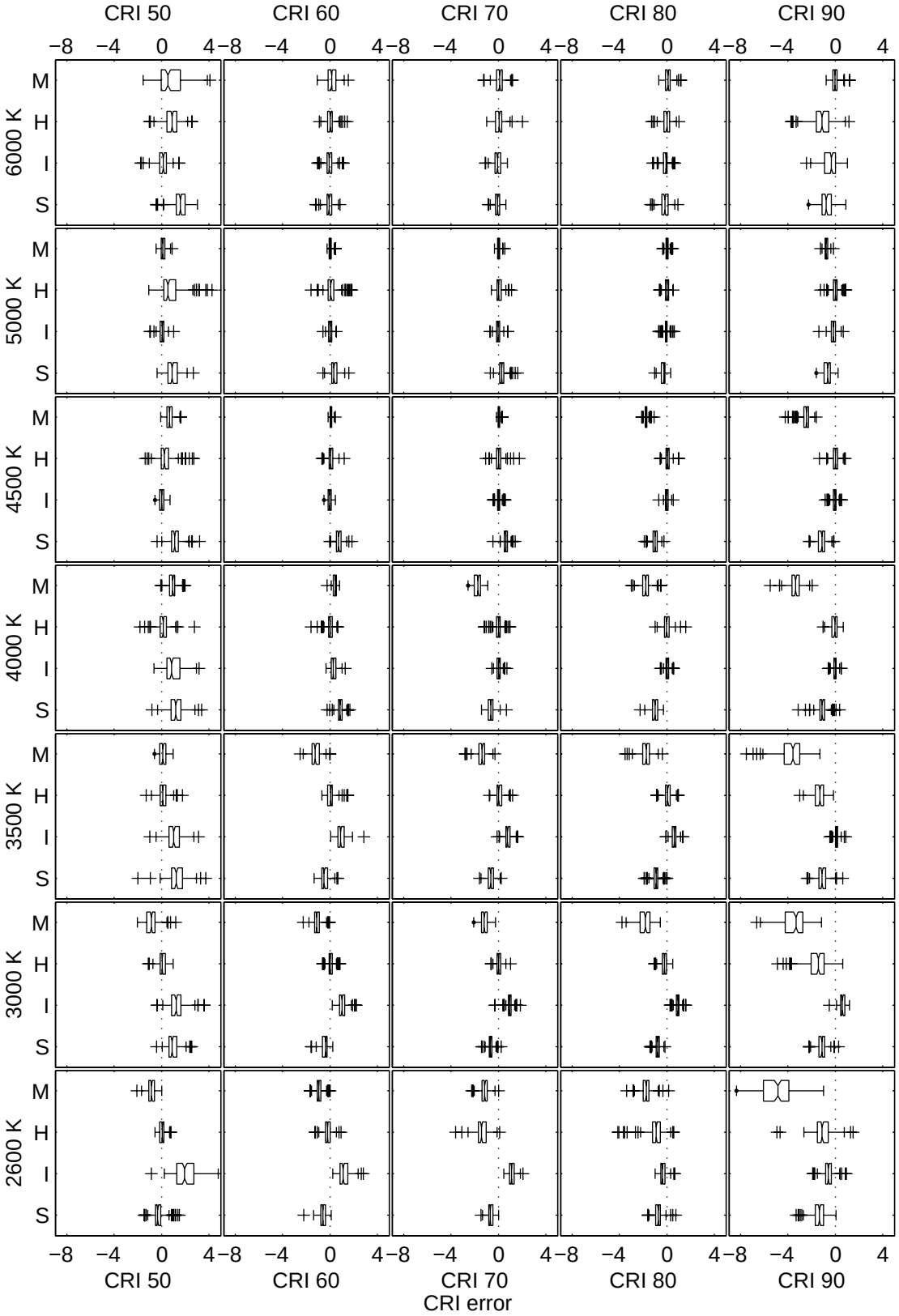

Figure 31. Boxplots of CRI error by color temperature and CRI, for all spectra (S, sulphur; I, incandescent; H, high pressure sodium; M, metal halide)

Figure 32 is a set of Boxplots of variability in efficiency (efficiency for each run minus median efficiency) for each $\overline{T}^c$ combination. The range of efficiency values resulting from this method (with the exception of the MH spectrum at low color temperatures and high CRI, discussed below) is quite small (Table 6).

Table 6. Range of values encompassing the center 50% (i.e., the interquartile range), 90%, and 98% of the variability in efficiency.

Spectrum	50%	90%	98%
Sulphur	0.005	0.015	0.027
Incandescent	0.004	0.013	0.029
High Pressure Sodium	0.002	0.016	0.045
Metal Halide	0.005	0.029	0.078

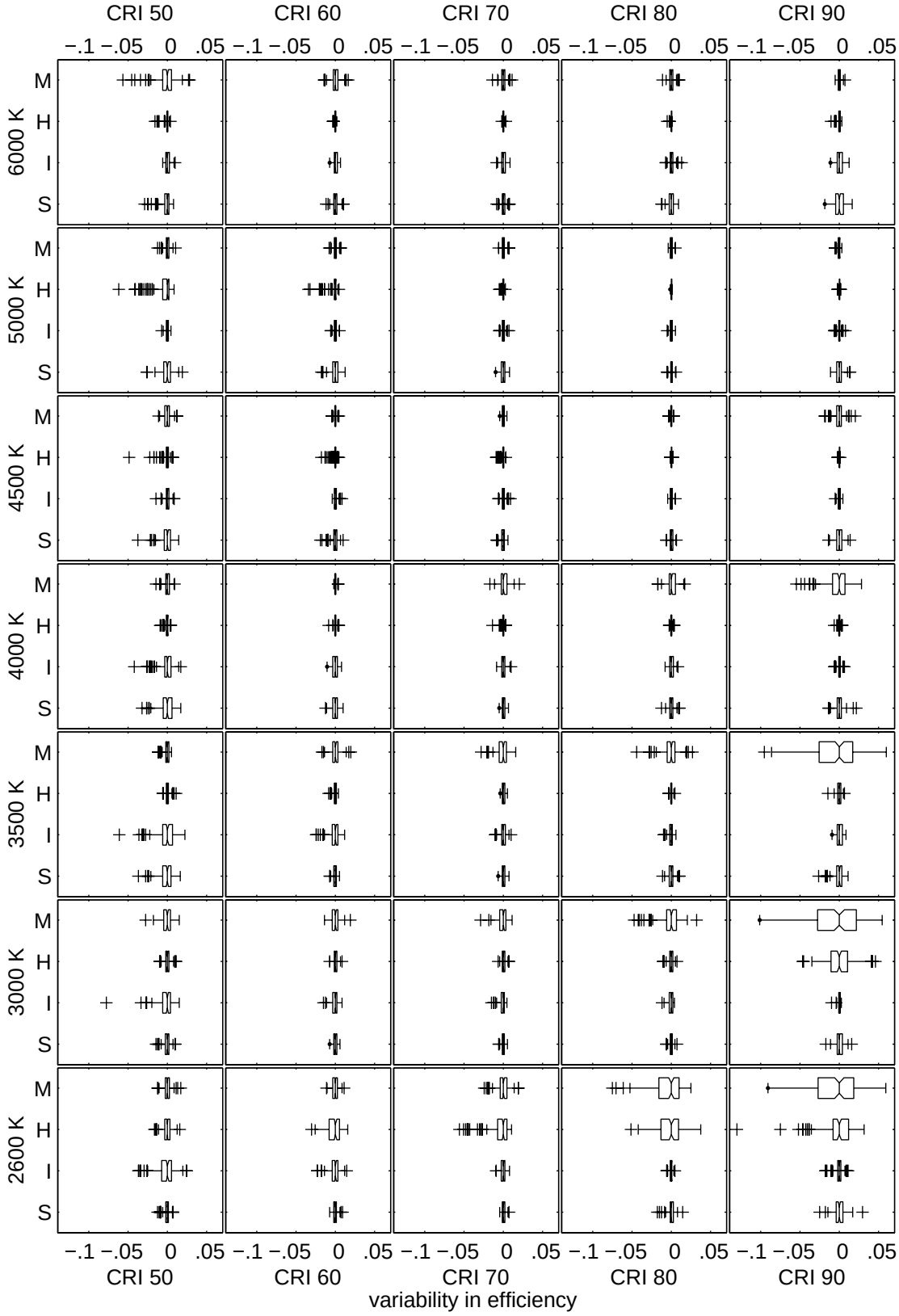

Figure 32. Boxplots of efficiency by color temperature and CRI, for all spectra (S, sulphur; I, incandescent; H, high pressure sodium; M, metal halide)

It was noted above that for the metal halide spectrum at low color temperature (below 3500 K) and high CRI (90), this method tends to produce filters with comparatively high CRI error and comparatively high variability in efficiency. The reason for this is likely idiosyncratic to the distribution of energy in the MH spectrum, which has a lot of energy around 570 nm (23% between 550 and 590 nm). The filters for the MH spectrum have a notch centered at 570 nm, which is wider for lower color temperatures (Appendix A, Figure 41). This results in a pronounced tradeoff between CRI and efficiency (small changes in CRI require comparatively large changes in efficiency). For example, Figure 33 is a plot of the relationship between CRI and efficiency at two different color temperatures; at the high color temperature, there is almost no relationship between CRI and efficiency (r^2 = 0.03), while at the lower color temperature, there is a clear relationship (r^2 = 0.63). Note that this is not a result of the way the fitness function is scaled – the filtered sulphur lamp spectrum has efficiency comparable to the filtered MH spectrum (Figure 34), but exhibits no significant correlation between CRI and efficiency.

Moving the MH spectrum to low color temperatures at high CRI is a comparatively difficult problem and suggests that if this were of particular interest, the GA parameters should be adjusted (e.g., increase the maximum generations, or adjust the scaling parameters for the $\overline{T}^c$ penalties) to compensate (note that the coefficients were tuned, by hand, to produce good performance overall, and not for individual spectra). However, for this application, it is not critical.

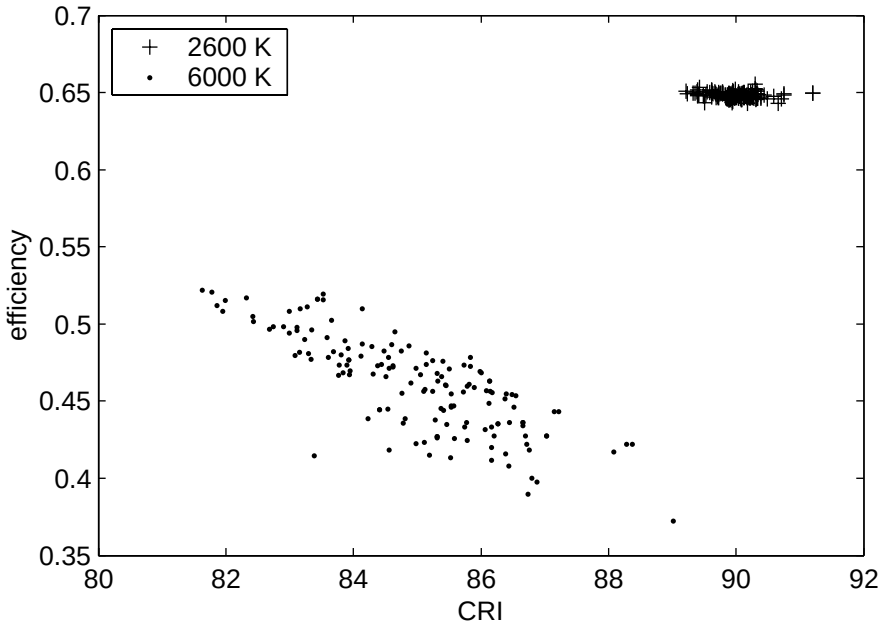

Figure 33. Relationship between efficiency and CRI at two different color temperatures. At the high color temperature, there is almost no relationship between CRI and efficiency (r^2 = 0.03), while at the lower color temperature, there is a clear relationship (r^2 = 0.63).

The results presented in this section suggest that the technique outlined here is a good approach for determining the tradeoff between chromaticity, CRI, and efficiency in filtered spectra. In particular, they show that this method results in solutions that are near the target chromaticity and CRI (both are well within industry tolerances), and that the variability in efficiency of the filtered spectra is small.

5.3.2 Lighting Results

Figure 34 is a series of plots of the tradeoff between chromaticity, CRI, and luminous efficacy (represented by the median efficiency across all runs) in filtered spectra. Note that not all of these points are Pareto-optimal: within each set of axes, only points to the right of the highest point are Pareto-optimal. For example, for the sulphur lamp at 3500 K, the first point (CRI 50) is not Pareto-optimal – it is dominated by the CRI 60 and 70

solutions (e.g., all other things being equal, why make a lamp at CRI 50 when one could make a lamp at CRI 70 with the same luminous efficacy).

This figure shows that there is a clear interaction between CRI and chromaticity with respect to efficiency (i.e., the effect of CRI on maximum attainable level of efficiency is dependant on the chromaticity, which was expected). For the incandescent and HPS spectra, over the range of both variables examined (i.e., the range of interest to industry), changes in chromaticity have a greater effect on maximum attainable efficiency than changes in CRI; for the sulphur and MH spectra, the effect of CRI and chromaticity is comparable.

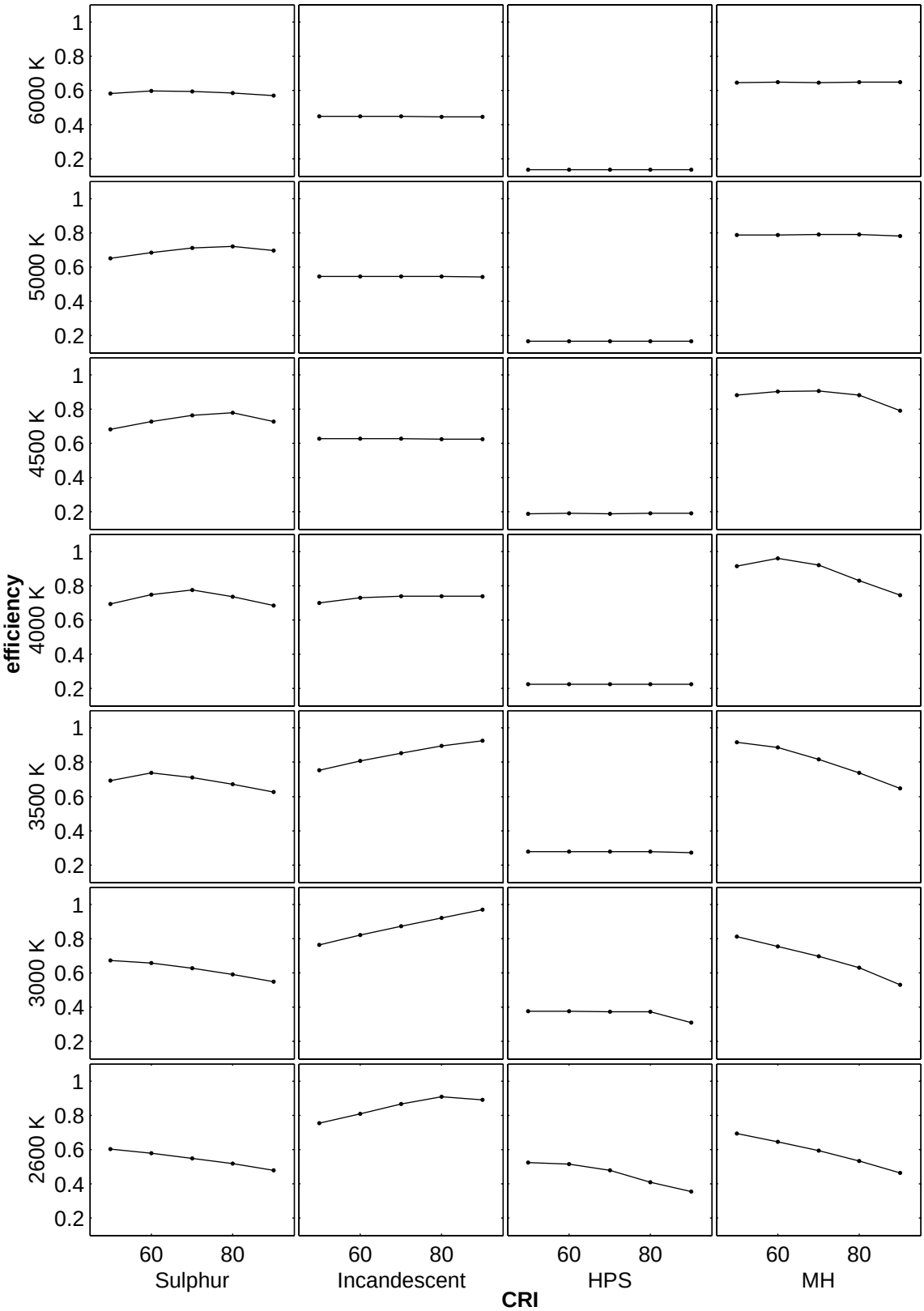

Figure 34. Median efficiency of filtered spectra for the four different lamp types. For each set of axes, the ordinate is efficiency and the abscissa is CRI.

Figure 35 is the same data, plotted a way that makes picking out Pareto-optimal points less easy, but, from the perspective of a lighting engineer, is more useful. The abscissa is color temperature, the ordinate is efficiency, and each solid curve is for a different CRI. Also plotted is the maximum efficiency (regardless of CRI) at each color temperature derived using MacAdam's (1935a) method. Note that (as discussed in the previous section) because the solutions found by the GA have some leeway in chromaticity, the efficiency of some the solutions found with the GA appear to be slightly greater than MacAdam's limit at some points (the GA "cheats" slightly, taking slight gains in efficiency for small error in chromaticity).

There are several interesting features highlighted by Figure 35. First, the maximum efficiency is quite close to the MacAdam limit for each chromaticity (again, with the exception of the metal halide lamp at low color temperatures). This is further evidence that the GA approach is finding true global optima: at *some* CRI, the efficiency for just chromaticity and for chromaticity and CRI optimized *must* be the same. Second, as color temperature is increased (i.e., the light becomes "more blue") the effect of CRI on maximum efficiency is decreased. The reason for this is that it is necessary (for these spectra) to remove a relatively broad band of energy in the yellow-red region of the spectrum to increase the color temperature; this permits a wider range of options for balancing CRI at about the same efficiency.

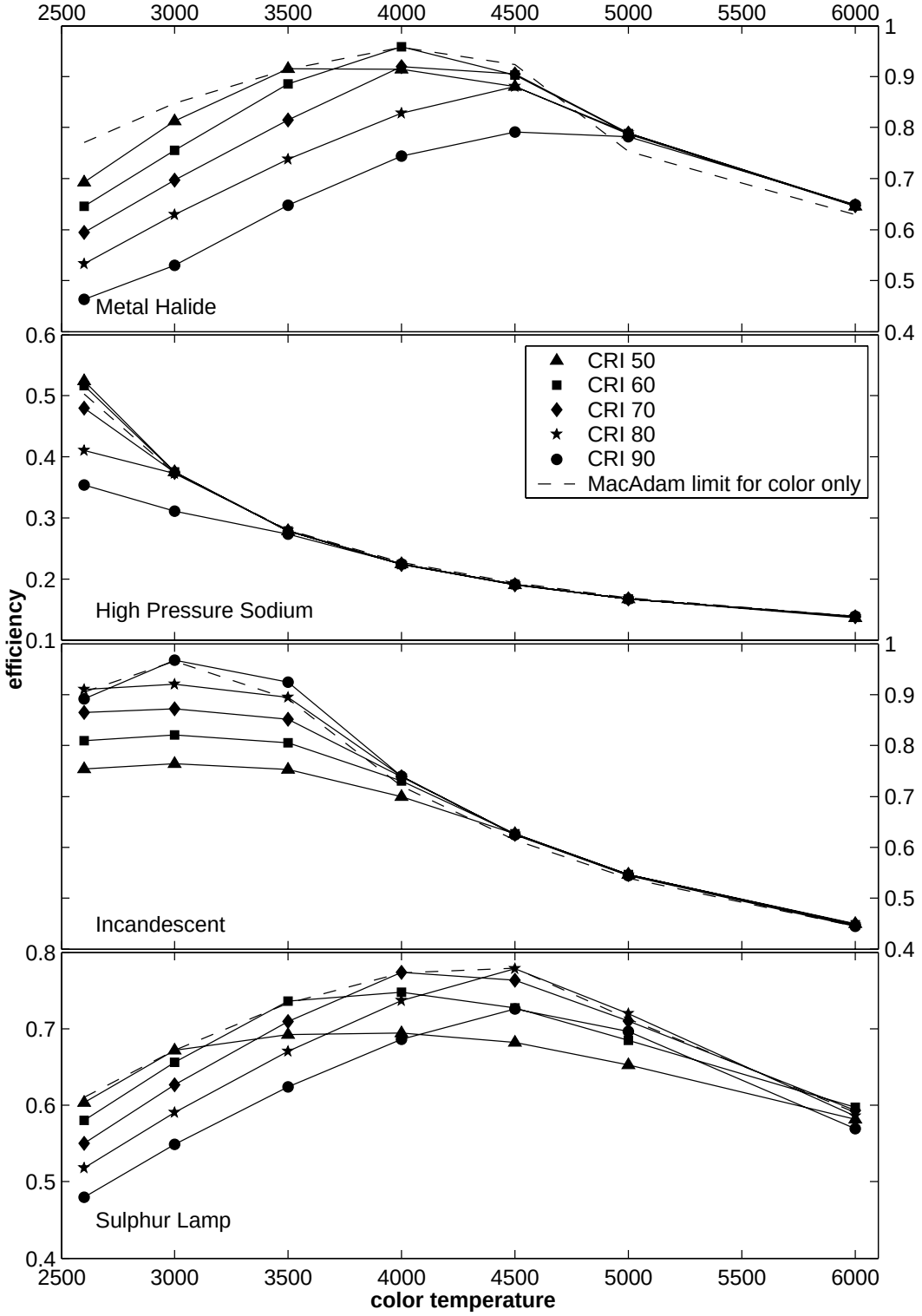

Figure 35. Tradeoff between efficiency and CRI at each chromaticity. Note that the maximum efficiency is quite close to the MacAdam limit for each chromaticity (except metal halide at low color temperatures). This is additional evidence that the GA approach is finding true global optima: at some CRI, the efficiency for just chromaticity and for chromaticity and CRI optimized must be the same. Also, as color temperature is increased the effect of CRI on maximum efficiency is decreased.

Figure 36 is a plot of the 50th, 25th and 75th percentiles of transmittance across all filters at each $\overline{T}^c$ for the incandescent spectrum (results from the other spectra are presented in Appendix A). Naturally, the filters evolved are different for the different spectra, but they do share some commonalities. For target x-chromaticities greater than the unfiltered lamp x-chromaticity (see Figures 25 and 26), some wavelengths shorter than 520 nm are invariably reduced; conversely, for target x-chromaticities less than the unfiltered lamp x-chromaticity, some wavelengths greater than 570 are attenuated. The incandescent spectrum (Figure 36) for 3000 K and 3500 K at CRI of 90 provides a good example of this: for 3500 K (which has a lower x chromaticity coordinate than the unfiltered spectrum), there is a substantial notch at 670 nm; for 3000 K (which has a higher x chromaticity coordinate than the unfiltered spectrum), there is a substantial notch at 490 nm.

In cases where a $\overline{T}^c$ is not on the Pareto-optimal front, the GA is forced to decrease color rendering at a minimum loss of efficiency, while holding chromaticity constant. For the smooth, continuous spectra (incandescent and sulphur), two widely separated notches appear to be the most efficient way to do this. The incandescent spectrum (Figure 36) at 2600 K is a good example of this. The CRIs of 90 and 80 are on the Pareto-optimal front (Figure 34); as CRI decreases below this, the short wavelength notch widens slightly, and a long wavelength notch grows substantially. For the sulphur and incandescent spectra, all Pareto-optimal filters have only one connected region of decreased transmittance. It is

difficult to make generalizations about the MH and HPS spectra, as the filters developed are substantially influenced by the idiosyncratic distribution of energy.

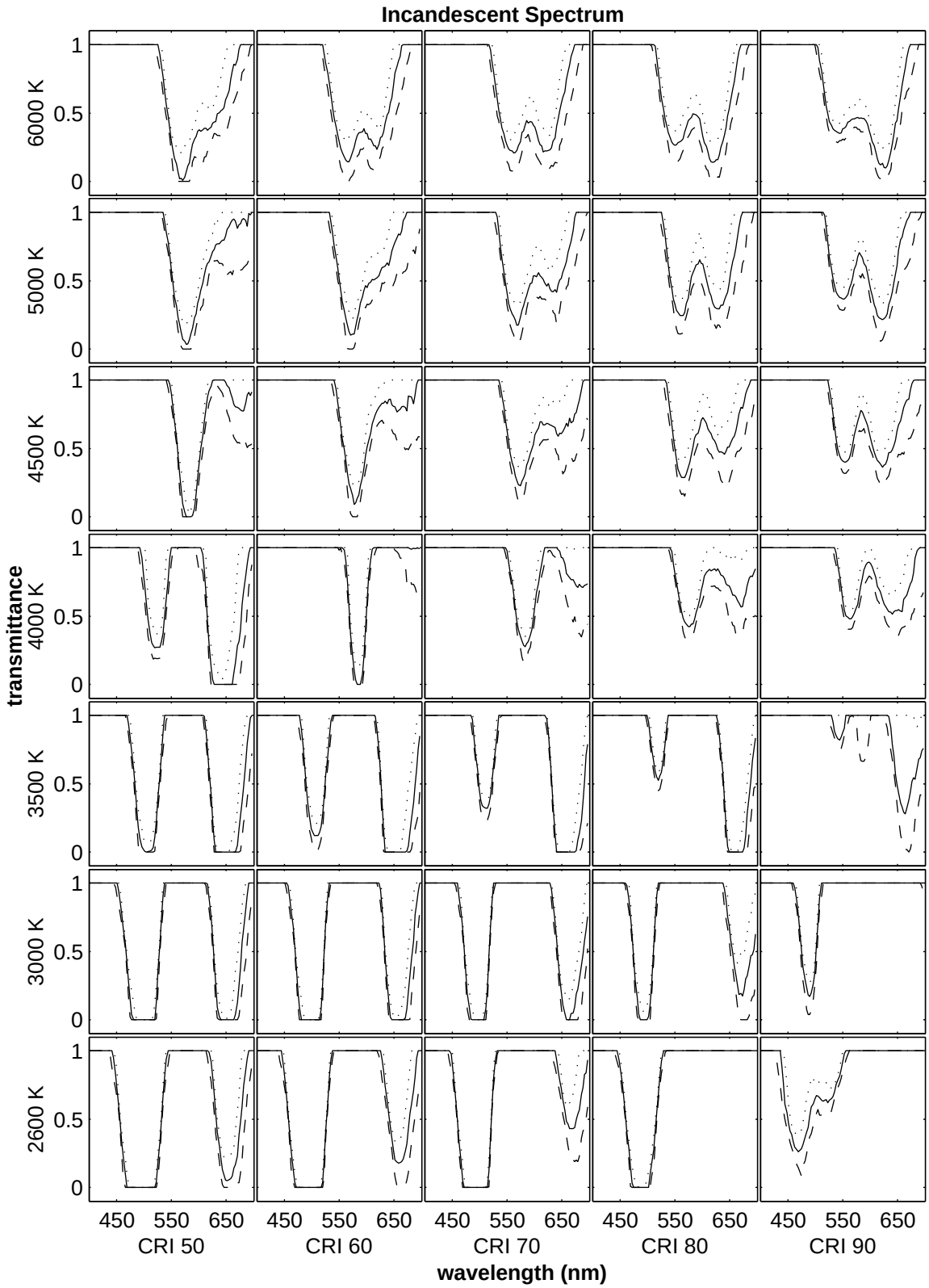

Figure 36. Median (solid line), 25th percentile (dashed line), and 75th percentile (dotted line) of the transmittance (at each wavelength sampled) of the filter designed for the incandescent spectrum, at each $\overline{T}^c$.

6. Conclusions and Future Work

While high luminous efficacy is an important property for new electric light source technologies, good colorimetric properties are required to be truly successful in the lighting marketplace. There are cases, e.g., the sulphur lamp, where the efficacy of a new electric light source technology is very high but the colorimetric properties are unsuitable for many applications. In cases like this, an effective strategy might be to filter the light source such that its colorimetric properties improve, but its efficacy remains competitive. The goal of this work was to develop a method to determine the tradeoff between luminous efficacy, apparent color, and color rendering index for filtered spectra.

By adopting a definition of photopic spectral luminous efficiency, the CIE defined the maximum attainable luminous efficacy (regardless of apparent color or CRI properties) for any lamp in 1924 (Wyszecki and Stiles, 1982). MacAdam (1935a, 1935b) determined how to filter an arbitrary SPD such that it achieves maximum efficiency for any apparent color, but did not account for CRI. Several studies[27] during the 1970s employ a variety of techniques to examine the relationship between efficacy, chromaticity, and CRI for line and band spectra. Thus, a technique for the simultaneous optimization of chromaticity, CRI, and luminous efficacy in filtered light sources with broadband spectra is a natural step in the evolution of lighting research.

[27] Koedam and Opstelten, 1971; Koedam, Opstelten, and Radielovic, 1972; Thornton, 1971; Einhorn and Einhorn, 1967; Walter, 1971; Haft and Thornton, 1972; Opstelten, Radielovic, and Verstegen, 1975; Walter, 1978

A GA approach to spectrum optimization for chromaticity and efficacy only was undertaken principally as a proof of concept, as MacAdam's work (1935a, 1935b) offers a covenant benchmark for comparison. The results produced using this method demonstrate two things. First, that the GA approach can produce filters that exhibit performance quite close to the optimal, but are quite different in form (e.g., Figure 16) from the (deterministically generated) optimum filter; this may have application, e.g., to providing alternatives for manufacturing[28]. Second, that the GA approach can develop solutions at or quite near known optima for this kind of problem (Figures 20-22), which suggests that GAs might reasonably be employed to solve the three objective (chromaticity, efficacy, and CRI) spectrum optimization problem, for which there is no general solution.

However, from a computational perspective, this approach is quite inefficient, insofar as during one run only one solution at a single chromaticity point is identified. Thus, for applications like determining the Pareto-optimal front for chromaticity and efficiency, which would require many separate runs of the algorithm, this approach is *extremely* inefficient, because good components of solutions ("building blocks") are not shared across the populations during separate runs, requiring the GA to reinvent the wheel each time. This suggested that an approach for developing solutions at many chromaticity points in parallel (and sharing good chunks of genetic material as these solutions are developed) was called for, which led to a novel approach to multiobjective optimization, the target objectives genetic algorithm (TOGA).

---

[28] Although if this were the goal, a slightly different approach would be called for, as this approach is not designed to insure variety.

TOGA is a non-Pareto, non-aggregating function approach to multiobjective optimization similar in concept to both goal programming and Schaffer's (1984, 1985) VEGA. TOGA was used to evolve 150 filters for each of four different lamps at 35 combinations of target chromaticity and CRI ($\overline{T}^c$). The properties of the filtered spectra were examined to assess the suitability of TOGA as a spectrum optimization procedure. The deviations from the target chromaticity and CRI were small (particularly when considering one set of coefficients was used for all spectra), and well within the limits of error commonly accepted by the lighting industry. The variability in efficiency of the filtered light sources was shown to be quite low, suggesting that the use of TOGA for optimizing filtered lamp spectra is a reliable technique.

TOGA has several advantages as an optimization method. First, the mechanics of the TOGA process do not require much computing power, particularly when compared to Pareto based approaches. TOGA takes advantage of *a priori* knowledge of the objective space to efficiently evolve solutions. By knowing where to search (for the example presented here, in chromaticity and CRI space), and driving the population toward these points, efficiency at these points of interest can be effectively maximized. The use of disparate target points for subpopulation formation insures diversity in the population overall, allows TOGA to generate multiple optimal points during each run, and allows these points to be generated on either a concave or convex Pareto front. Moreover, because TOGA shares good building blocks developed during a run throughout a population , substantial computing time is saved when compared to individual runs of an

aggregating GA approach. For example, if the approach employed in Chapter 4 (i.e., individual runs for each target combination) were employed to examine the target values listed in Table 3, 875,000 chromosomes would have to be evaluated[29]; TOGA can generate the same results in 87,500 chromosome evaluations[30] (note that this argument is independent of chromosome encoding for a fixed number of generations).

However, TOGA also has a several limitations. TOGA is most efficient when the researcher has some domain knowledge and is able to select good combinations of objectives, and the resulting solutions (as with goal programming) may not be Pareto-optimal, depending on the target points chosen, but are optimal solutions at that particular combination of objectives. Another limitation is that some experimentation is typically required to determine good scaling factors for the objectives.

Finally, it is important to note that while TOGA was employed to solve a lighting problem here, it is in principal a general purpose methodology, and might be used to solve any multiobjective optimization problem.

6.1 Practical Application of this Method

The core problem in this dissertation was inspired by a real-world engineering problem. The sulphur lamp was developed in the late 1980s, and was commercialized during the early to mid 1990s. The sulphur lamp has a very high luminous efficacy (around 100 lm/W), a color rendering index of 78, and a greenish-white apparent color. The sulphur lamp was found to be a good source for industrial applications (e.g., Eklund and Boyce,

[29] 35 runs * 50 chromosomes/run * 500 generations
[30] 1 run * 175 chromosomes * 500 generations

1997a, 1997b, 1997c; Richman, Heerwagen, and Hollomon, 1998), and for lighting public spaces where color is not important (e.g., subway platforms; Boyce and Eklund, 1997). However, the apparent color of the sulphur lamp limited its application, and the company manufacturing it ultimately withdrew it from the market while it sought ways to produce better apparent color (through filtering and other means).

The method outlined in this dissertation shows that the sulphur lamp is capable of being filtered to a color temperature of 4000 K or 4500 K, at a CRI of 70 or 80 (respectively), at a luminous efficacy of around 80 lm/W. Given the many other desirable properties (high efficacy, very long lamp life, stable apparent color when dimmed, stable luminous efficacy when dimmed, small size, stable color over the life of the lamp, etc.) of the sulphur lamp technology, such filtered lamps would find many applications where colorimetric properties are important. For example, the filtered sulphur lamp described above would probably compete well (depending on price) with metal halide lamps, which have similar chromaticity and luminous efficacy, but lower CRI, shorter lamp life, poor luminous efficacy when dimmed, and unstable color over the life of the lamp.

As other light source technologies are developed (particularly broad spectrum sources) with undesirable colorimetric properties (e.g., white light emitting diodes), this methodology can be applied to determine how the lamp might be filtered to result in a more marketable lamp. The most efficient way to do this is in two stages. The first stage is to use MacAdam's (1935a) method (which is very fast) to test the variety of spectra that can typically be produced using the novel lamp technology (e.g., by varying the

physical parameters of the lamp, such as pressure, chemical composition, temperature, etc.). This will establish an absolute upper bound for filtered efficacy, and can show, at minimal computational cost, what combination of physical parameters might be competitive when filtered (note that it might be practical to manufacture two versions of a lamp, one meant to be used unfiltered for industrial applications, and the other to be filtered for non-industrial applications). However, good apparent color and high efficiency are not enough to be a marketable light source; it must also have good color rendering. Therefore, having identified promising spectra during the first stage, the second stage is to use the method described in this dissertation to determine the tradeoff between CRI, efficacy, and chromaticity. Given these results and other known properties of the lamp (e.g., source life, source size, stability and efficacy when dimmed, etc.) the marketability of the filtered lamp in comparison to other currently used light sources can be assessed.

6.2 Future Work

There are several opportunities for future work. It would be interesting to assess the performance of TOGA on other multiobjective optimization problems, and to compare its performance on these problems to other multiobjective optimization methods.

An interesting approach would be to hybridize the TOGA method with traditional optimization methods – the GA would be used to get close to the global optimum (avoiding local optima) and the traditional optimization method (e.g., the Nelder-Mead simplex method; Nelder and Mead, 1965) could be used as a starting point to find the exact optimum. Goel and Deb (2001) tested a similar approach, and found that, "…hybrid

methods are candidates of being a good and robust algorithm for solving the real-world multiobjective optimization problems." [sic].

An interesting extension of this method would be to use the eight individual values of the CIE Special Color Rendering Index, R_i, as individual objectives. For this implementation, it would make more sense to treat R_i as thresholds rather than targets. This would overcome one common criticism of the general CRI, that occasionally one reference color will be significantly skewed but the overall CRI will remain high (which may be the case, given the shape of some of the filters developed using TOGA). The addition of this constraint would add non-trivial value for the lighting industry.

There are numerous other areas where the techniques outlined here might be applied. For example, rather than luminous efficacy, many safety related applications (e.g., signal lights) are concerned with maximizing conspicuity. Researchers interested in roadway lighting might be interested in optimizing lamps for both foveal and off-axis sensitivity (e.g., determining the tradeoff between photopic luminous efficacy and some measure of mesopic luminous efficacy). The chromosome encoding and related specialized mutation methods outlined in Chapter 4 might be easily adapted to electronic filtering applications.

7 References

Aherne, F., Thacker, N., and Rockett, P. (1997). Optimal pairwise geometric histograms. In Clark (ed.), *Electronic Proceedings of the Eighth British Machine Vision Conference*, available at http://www.bmva.ac.uk/bmvc/1997/papers/071/bmvc.html [accessed August 2002].

Allenson, R. (1992). *Genetic Algorithms With Gender for Multi-Function Optimization*. Technical Report EPCC-SS92-01, Edinburgh Parallel Computing Centre, Edinburgh, Scotland.

Alman, D. (1993). Industrial color-difference evaluation progress report (CIE Technical Committee 1-29). *Color Research and Application*, 18, 137.

ANSI (1996). *American National Standard for Electric Lamps: Specifications for the chromaticity of Fluorescent Lamps*. American National Standards Institute (ANSI), ANSI C78.376-1996.

Ashdown, I. (1994). Non-imaging optics design using genetic algorithms. *Journal of the Illuminating Engineering Society*. 23, 12-21.

Belegundu, A., Murthy, D., Salagame, R., and Constants, E. (1994). Multiobjective optimization of laminated ceramic composites using genetic algorithms. In *Fifth AIAA/USAF/NASA Symposium on Multidisciplinary Analysis and Optimization*, American Institute of Aeronautics and Astronautics, Reston, VA, 1015-1022.

Borbely, A., Samson, A., and Schanda, J. (2001). The concept of correlated colour temperature revisited. *Color Research and Application*, 26, 450-457.

Boyce, P., and Eklund, N. (1997). *An Evaluation of Subway Platform Lighting Using the Solar 1000 Sulphur Lamp.* Technical Report, Lighting Research Center, Troy, NY.

Brown, W. (1952). The effect of field size and chromatic surroundings on color discrimination. *Journal of the Optical Society of America*, 42, 837-844.

Brown, W. (1957). Color discrimination of twelve observers. *Journal of the Optical Society of America*, 47, 137-143.

Brown, W., Howe, W., Jackson, J., and Morris, R. (1956). Multivariate normality of the color-matching process. *Journal of the Optical Society of America*, 46, 46-49.

Brown, W., and MacAdam, D. (1949). Visual sensitivities to combined chromaticity and luminance differences. *Journal of the Optical Society of America*, 39, 808-834.

CIE (1971). *Colorimetry (Official Recommendations of the International Commission on Illumination)*. CIE Publication No. 15 (E-1.3.1), Bureau Central de la CIE, Paris.

CIE (1995a). *Industrial Color Difference Evaluation*. CIE Technical Report 116, Commission International de L'Eclairage, Vienna.

CIE (1995b). *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*. CIE Publication No. 13.3-1995, Bureau Central de la CIE, Paris.

Carleton, S., Seinen, P., and Stoffels, J. (1997). Metal halide lamps with ceramic envelopes: A breakthrough in color control. *Journal of the Illuminating Engineering Society*. 26, 139-145.

Charnes, A., Cooper, W., and Ferguson, R. (1955) Optimal estimation of executive compensation by linear programming. *Management Science*, 1, 138-151.

Chipperfield, A., and Fleming, P. (1995). Gas turbine engine controller design using multiobjective genetic algorithms. In Zalzala (ed.), *Proceedings of the First IEE/IEEE International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications*, Inspec/IEE, 214-219.

Coello, C. (1996). *An Empirical Study of Evolutionary Techniques for Multiobjective Optimization in Engineering Design*. PhD thesis, Department of Computer Science, Tulane University, New Orleans, LA.

Coello, C. (1999). A comprehensive survey of evolutionary-based multiobjective optimization techniques. *Knowledge and Information Systems*, 1, 129-156.

Coello, C. (2000). Handling preferences in evolutionary multiobjective optimization: A survey. In *Congress on Evolutionary Computation* (vol 1), IEEE Service Center, New York, 30-37.

Coello C., and Christiansen, A. (1998). Two new GA-based methods for multiobjective optimization. *Civil Engineering Systems*, 15, 207-243.

Coello, C. and Christiansen, A. (1999). MOSES: A multiobjective optimization tool for engineering design. *Engineering Optimization*, 31, 337-368.

Coello C., Christiansen, A., and Hernandez A. (1998). Using a new GA-based multiobjective optimization technique for the design of robot arms. *Robotica*, 16, 401-414.

Coello, C., Hernandez, F., and Farrera, F. (1997). Optimal design of reinforced concrete beams using genetic algorithms. *Expert Systems with Applications: An International Journal*, 12, 101-108.

Coello, C., van Veldhuizen, D., and Lamont, G. (2002). *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic, New York.

Cohon, J. (1978). *Multiobjective Programming and Planning*. Academic Press, New York.

COST (2001). *Memorandum of Understanding for the implementation of a European Concerted Research Action designated as COST Action 529 "Efficient Lighting for the 21st Century"*. European Co-operation in the field of Scientific and Technical Research, available at http://www.efficient-lighting.org/ [accessed July 2002].

Cvetković, D., Parmee, I., and Webb, E. (1998). Multiobjective optimization and preliminary airframe design. In Parmee (ed), *Adaptive Computing in Design and Manufacture: The Integration of Evolutionary and Adaptive Computing Technologies with Product/System Design and Realization*, Springer Verlag, Berlin, 255-267.

Darwin, C. (1896). *The Origin of Species by Means of Natural Selection, or, The Preservation of Favored Races in The Struggle for Life*. 7th edition, D. Appleton & Co., New York.

Deb, K. (1998a). *Multi-Objective Genetic Algorithms: Problem Difficulties and Construction of Test Problems*. Technical Report CI-49/98, Department of Computer Science, University of Dortmund, Germany.

Deb, K. (1998b). *Non-linear Goal Programming Using Multi-Objective Genetic Algorithm*. Technical Report CI-60/98, Department of Computer Science, University of Dortmund, Germany.

Deb, K. (1999). Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation*, 7, 205-230.

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, West Sussex, England.

Deb, K., and Kumar, A. (1995). Real-coded genetic algorithms with simulated binary crossover: Studies on multimodal and multiobjective problems. *Complex Systems*, 9, 431-454.

Denbigh, P., and Jones, B. (1984). A simple computer program for designing standard high-pressure sodium lamps. *Lighting Research & Technology*. 16(4), 193-199.

Duckstein, L. (1984). Multiobjective optimization in structural design: The model choice problem. In Atrek, Gallagher, Ragsdell, and Zienkiewicz (eds.), *New Directions in Optimum Structural Design*, John Wiley and Sons, New York, 459-481.

Dupont, D. (2002). Study of the reconstruction of reflectance curves based on tristimulus values: Comparison of methods of optimization. *Color Research and Application*, 27, 88-99.

Einhorn, H. and Einhorn, F. (1967). Inherent efficiency and color rendering of white light sources. *Illuminating Engineering*, 62, 154-158.

Eklund, N., and Boyce, P. (1997a). *A Comparison of a Solar 1000 Sulphur Lamp Installation and Conventional Fluorescent Installations in Two Similar Post Terminals*. Technical Report, Lighting Research Center, Troy, NY.

Eklund, N., and Boyce, P. (1997b). *An Evaluation of the Lighting of an Airport Apron Lit by the Solar 1000 Sulphur Lamp*. Technical Report, Lighting Research Center, Troy, NY.

Eklund, N., and Boyce, P. (1997c). *An Evaluation of the Lighting of a Parcel Sorting Terminal Employing the Solar 1000 Sulphur Lamp*. Technical Report, Lighting Research Center, Troy, NY.

Estevez, O. (1982). A better colorimetric standard observer for color vision studies: The Stiles and Burch 2° color matching functions. *Color Research and Application*, 7, 131-134.

Fairchild, M. (1998). *Color Appearance Models*. Addison Wesley, Reading, Massachusetts, USA

Fairman, H., Brill, M., and Hemmendinger, H. (1997). How the CIE 1931 color-matching functions were derived from Wright-Guild data. *Color Research and Application*, 22, 11-22.

Fonseca C., and Fleming, P. (1993). Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. In Forrest (ed), *Proceedings of the Fifth International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo, CA, 416-423.

Fonseca, C., and Fleming, P. (1995). An overview of evolutionary algorithms in multiobjective optimization, *Evolutionary Computation*, 3, 1-16.

Goel, T., and Deb, K. (2001). *Hybrid Methods for Multi-Objective Evolutionary Algorithms*. KanGAL Report Number 2001004, Kanpur Genetic Algorithms Laboratory, Indian Institute of Technology, Kanpur, India.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison Wesley Longman, Inc., Reading, Massachusetts.

Goldberg, D. (1990). *Real-coded genetic algorithms, virtual alphabets, and blocking*. IlliGAL Report 90001, Illinois Genetic Algorithms Laboratory, Dept. of General Engineering, University of Illinois, Urbana, IL.

Goldberg, D. (2002). *The Design of Innovation: Lessons from and for Competent Genetic Algorithms*. Kluwer Academic Publishers, Norwell, MA.

Goldberg, D., and Deb, K. (1991). A comparison of selection schemes used in genetic algorithms. In *Foundations of Genetic Algorithms*, Rawlins, G. (Ed.), 69. M. Kaufmann Publishers, San Mateo, Calif, 69-93.

Goldberg, D., and Richardson, J. (1987). Genetic algorithms with sharing for multimodal function optimization. In Grefenstette (ed.), *Proceedings of the Second International Conference on Genetic Algorithms*, Lawrence Erlbaum Associates, Hillsdale, NJ, 41-49.

Green, F. (1994). A method for utilizing diploid/dominance in genetic search. In, Michalewicz, Z. Schaffer, J., Schwefel, H., Fogel, D., and Kitano, H. (Eds.), *Proceedings of the First IEEE International Conference on Evolutionary Computation*, v1, 439. IEEE Service Center, Piscataway, NJ, 171-176.

Grefenstette, J. (1984). GENESIS: A system for using genetic search procedures. In *Proceedings of the 1984 Conference on Intelligent Systems and Machines*, 161-165.

Grefenstette, J., (ed.) (1985). *Procedings of the First International Conference on Genetic Algorithms and Their Applications*. Lawrence Erlbaum, Hillsdale, NJ.

Gu, G., Burrows, P. E., Venkatesh, S., (1997). Vacuum-deposited, nonpolymeric flexible organic light-emitting devices, *Optics Letters*, 22, 172-175,.

Guild, J. (1931). The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society, A*, 230, 149-187.

Haft, H., and Thornton, W. (1972). High performance fluorescent lamps. *Journal of the Illuminating Engineering Society*. 1, 29-35.

Haimes, Y., Hall, W., and Freedman, H. (1975). *Multiobjective Optimization in Water Resources Systems*. Elsevier Scientific, Amsterdam.

Hajela P., and Lin C., (1992). Genetic search strategies in multicriterion optimal design. *Structural Optimization*, 4, 99-107.

Halliday, D. and Resnick, R (1986). *Physics* (part two, extended version, 3rd. Ed.). John Wiley & Sons, New York.

Haupt, R. and Haupt, S. (1998). *Practical Genetic Algorithms*. Wiley-Interscience, New York.

Hilliard, M., Liepins, G., Palmer, M., and Rangarajen, G. (1989). The computer as a partner in algorithmic design: Automated discovery of parameters for a multiobjective scheduling heuristic. In Sharda, Golden, Wasil, Balci, and Stewart (eds.), *Impacts of Recent Computer Advances on Operations Research*, North-Holland Publishing Company, New York.

Holland, J. (1962). Outline for a logical theory of adaptive systems. *Journal of the Association of Computing Machinery*, 9, 297-314.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, Michigan.

Horn, J., and Nafpliotis, N. (1993). *Multiobjective Optimization using the Niched Pareto Genetic Algorithm*. Technical Report, IlliGAl Report 93005, University of Illinois at Urbana-Champaign, Urbana, IL.

Horn, J., Nafpliotis, N. and Goldberg, D. (1994). A niched Pareto genetic algorithm for multiobjective optimization. *Proceedings of the first IEEE conference on evolutionary computation*, IEEE Service Center, Piscataway, NJ, 82-87.

Houck, C., Joines, J., and Kay, M. (1995). *A Genetic Algorithm for Function Optimization: A MATLAB Implementation*. North Carolina State University - Industrial Engineering Technical Report 95-09.

IESNA (2000). *The IESNA Lighting Handbook* (9th Ed.). Illuminating Engineering Society of North America, New York.

Ignizio, J. (1976). *Goal Programming and Extensions*. Heath, Lexington, MA.

Ignizio, J. (1981). The determination of a subset of efficient solutions via goal programming. *Computing and Operations Research*, 3, 9-16.

Ivey, H (1963). Color and efficiency of luminescent light sources. *Journal of the Optical Society of America*, 53, 1185-1197.

Jakob, W., Gorges-Schleuter, M., and Blume, C. (1992). Application of genetic algorithms to task planning and learning. In Manner and Manderick (eds.), *Parallel Problem Solving From Nature*, (2nd Workshop), Lecture Notes in Computer Science, North-Holland Publishing Company, Amsterdam, 291-300.

Jones, G., Brown, R., Clark, D., Willett, P., and Glen, R. (1993). Searching databases of two-dimensional and three-dimensional chemical structures using genetic algorithms. In Forrest (ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms,* Morgan Kaufmann, San Mateo, CA, 597-602.

Judd, D (1933). The 1931 ICI standard observer and coordinate system for colorimetry. *Journal of the Optical Society of America*, 23, 359-372.

Judd, D. (1936). Estimation of chromaticity differences and nearest color temperature on the standard 1931 ICI colorimetric coordinate system. *Journal of the Optical Society of America*, 26, 421-425.

Judd, D. (1951). Report of U. S. secretariat committee on colorimetry and artificial daylight. In *CIE Proceedings*, v 1, part 7, p. 11 (Stockholm, 1951), Bureau Central de la CIE, Paris.

Judd, D. (1961). Maxwell and modern colorimetry. *The Journal of Photographic Science*, 9, 341-352.

Judd, D. and Wyszecki, G. (1975). *Color in Business, Science, and Industry* (3rd. Ed.). John Wiley & Sons, New York.

Kaiser, P., and Boynton, R. (1996) *Human Color Vision* (2nd Ed). Optical Society of America, Washington, DC.

Kandel, G., Eklund, N., and Schroeder, J. (1992). On the possibility of visually significant intraocular photoluminescence. In *Advances in Color Vision* (1992 Technical Digest Series volume 4), Optical Society of America, Washington, DC, 143-145.

Kandel, G., Schroeder, J., Geri, G., Fleming, K., Silvestri, M., and Eklund, N. (1993). A failed attempt to measure, in vivo, preretinal absorption. *Investigative Ophthalmology & Visual Science* (Supplement), 3, 776.

Koedam, M., and Opstelten, J. (1971). Measurement and computer-aided optimization of spectral power distributions. *Lighting Research & Technology*. 3(3), 205-210.

Koedam, M., Opstelten, J., and Radielović, D. (1972). The application of simulated spectral power distributions in lamp development. *Journal of the Illuminating Engineering Society*. 1, 285-289.

Koza, J. (1995). Survey of genetic algorithms and genetic programming. In *Proceedings of the Wescon 95 - Conference Record: Microelectronics, Communications Technology, Producing Quality Products, Mobile and Portable Power, Emerging Technologies*. IEEE, New York, NY, 589-594.

Krasko, Z., Brates, N., and Nortrup, E. (1998). Optimization of metal halide lamp design for better color rendering and color control. *Journal of the Illuminating Engineering Society*. 27, 152-159.

Liepins, G., Hilliard, M., Richardson, J., and Palmer, M. (1990). Genetic algorithms application to set covering and travelling salesman problems. In Brown and White (eds.), *Operations research and Artificial Intelligence: The integration of problem-solving strategies*, Kluwer Academic, Norwell, MA.

Lis, J., and Eiben, A. (1996). A multi-sexual genetic algorithm for multiobjective optimization. In Fukuda and Furuhashi (eds.), *Proceedings of the 1996 International Conference on Evolutionary Computation*, IEEE, 59-64.

Liu, X., Begg, D., and Fishwick, R. (1998). Genetic approach to optimal topology/controller design of adaptive structures. *International Journal for Numerical Methods in Engineering*, 41, 815-830.

MacAdam, D. (1935a). The theory of the maximum visual efficiency of colored materials. *Journal of the Optical Society of America*, 25, 249-252.

MacAdam, D. (1935b). Maximum visual efficiency of colored materials. *Journal of the Optical Society of America*, 25, 361-367.

MacAdam, D. (1937). Projective transformations of I.C.I. color specifications. *Journal of the Optical Society of America*, 27, 294-299.

MacAdam, D. (1942). Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America*, 32, 247-273.

MacAdam, D. (1943). Specification of small chromaticity differences. *Journal of the Optical Society of America*, 33, 18-26.

MacAdam, D. (1950). Maximum attainable luminous efficiency of various Chromaticities. *Journal of the Optical Society of America*, 40, 120.

Mäntyjärvi, M., and Tuppurainen, K. (1996). Color vision in patients with a silicone intraocular lens. *Journal of Cataract and Refractive Surgery*, 22 (supplement 2), 1308-1312.

Maxwell, J. (1872). On colour vision. *Proceedings of the Royal Institution of Great Britain*, 6, 260-271.

Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs* (3rd. Ed). Springer-Verlag, Berlin.

Michalewicz, Z. and Fogel, D. (2000). *How to Solve It: Modern Heuristics*. Springer-Verlag, Berlin.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, Massachusetts.

Narendran, N., Vasconez, S., Boyce, P., and Eklund, N. (2000). Just-perceivable color difference between similar light sources in display lighting applications. *Journal of the Illuminating Engineering Society*, 29, 68-77.

Nelder, J., and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7, 308-313.

Ohta, N. and Wyszecki, G. (1976a). Color changes caused by specified changes in the illuminant. *Color Research and Application*, 1, 17-21.

Ohta, N. and Wyszecki, G. (1976b). Designing illuminants that render given objects in prescribed colors. *Journal of the Optical Society of America*, 66, 269-275.

Opstelten, J., Radielović, D., and Verstegen, J. (1975). Optimum spectra for light sources. *Philips Technical Review*, 35, 361-370.

OSI (2002). *What are MacAdam Ellipses or color ovals?* Osram Sylvania Inc. Frequently Asked Questions, available at: http://www.sylvania.com/forum/pdfs/faq0026-0999.pdf [accessed July 2002].

Osyczka, A. (1978). An approach to multicriterial optimization problems for engineering design. *Computer Methods in Applied Mechanics and Engineering*, 15, 309-333.

Poloni, C. and Pediroda, V. (1997). GA coupled with computationally expensive simulations: tools to improve efficiency. In Quagliarella, Periaux, Poloni, and Winter (eds.), *Genetic Algorithms and Evolution Strategies in Engineering and Computer Science*, John Wiley & Sons, West Sussex, UK, 267-288.

Quagliarella, D. and Vicini A. (1997). Coupling genetic algorithms and gradient based optimization techniques. In Quagliarella, Periaux, Poloni, and Winter, (eds.), *Genetic Algorithms and Evolution Strategies in Engineering and Computer Science: Recent Advances and Industrial Applications*. John Wiley and Sons, West Sussex, UK, 289-309.

Ranjithan, S., Eheart, J., and Liebman, J. (1992). Incorporating fixed cost component of pumping into stochastic groundwater management: A genetic algorithm based optimization approach. *Eos Transactions*, American Geophysical Union, spring meeting supplement, 73,125.

Rao, S. (1986). Game theory approach for multiobjective structural optimization. *Computers and Structures*, 25, 119-127.

Richardson, J., Palmer, M., Liepins, G., and Hilliard, M. (1989). Some guidelines for genetic algorithms with penalty functions. In Schaffer (ed), *Proceedings of the Third International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo, CA, 191-197.

Richman, E., Heerwagen, J., and Hollomon, J. (1998). *Demonstration and Assessment of a Sulfur Lamp Retrofit Lighting System at Hill Air Force Base, Utah*. Technical Report, Pacific Northwest National Laboratory (Buildings Program), Richland, WA.

Ritzel, B., Eheart, J., and Ranjithan, S. (1994). Using genetic algorithms to solve a multiple objective groundwater pollution containment problem. *Water Resources Research*, 30, 1589-1603.

Robertson, A. (1968). Computation of correlated color temperature and distribution temperature. *Journal of the Optical Society of America*, 58, 1528-1535.

Said, F., and Weale, R. (1959). The Variation with age of the spectral transmissivity of the living human crystalline lens. *Gerontologia*, 3, 213-231.

Sandgren, E. (1994). Multicriteria design optimization by goal programming. In Adeli (ed.), *Advances in Design Optimization*, Chapman & Hall, London, 225-265.

Schaffer, J. (1984). *Multiple objective optimization with vector evaluated genetic algorithms*. Ph.D. thesis, Vanderbilt University, Nashville, TN.

Schaffer, J. (1985). Multiple objective optimization with vector evaluated genetic algorithms. In Grefenstette (Ed.), *Procedings of the First International Conference on Genetic Algorithms and Their Applications*. Lawrence Erlbaum, Hillsdale, NJ, 93-100.

Shaw, M. (1997). *Evaluating the 1931 CIE Color Matching Functions*. Masters Thesis, Graphic Media Studies, University of Hertfordshire, England.

Silk, S. (1999). *Lens Autofluorescence: In Aging and Cataractous Human Lenses, Clinical Applicability*. Doctoral Dissertation, Department of Ophthalmology, University of Oulu, Oulu, Finland.

Siminovitch, M., Gould, C., and Page, E. (1997). A high-efficiency indirect lighting system utilizing the solar 1000 sulfur lamp. *Proccedings of the Right Light 4 Conference*, Nov. 19-21, Copenhagen, Denmark.

Smith, V., and Pokorny, J. (1975). Spectral sensitivity of the foveal cone pigments between 400 and 500 nm. *Vision Research*, 15, 161-171.

Spears, G. (1974). Spectroradiometry photometry. *Journal of the Illuminating Engineering Society*. 3, 229-235.

Spears, W., De Jong, Bäck, K., Fogel, D. and de Garis, H. (1993). An overview of evolutionary computation. In Brazdil, P (ed.), *Proceedings of the European Conference on Machine Learning*. Springer Verlag, Vienna, Austria, 442-459.

Srinivas, N. and Deb, K. (1995). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2, 221-248.

Stanley, T., and Mudge, T. (1995). A parallel genetic algorithm for multiobjective microprocessor design. In Eshelman (ed.), *Proceedings of the Sixth International Conference on Genetic Algorithms*, Morgan Kaufmann Publishers, San Mateo, CA, 597-604.

Stiles, W., and Burch, J., (1955). Interim report on the Commission Internationale de l'Eclairage, Zurich, on the National Physical Laboratory's investigation of colour matching. *Optica Acta*, 2, 168-181.

Stockman A., and Sharpe, L. (1998). Human cone spectral sensitivities: a progress report. *Vision Research*, 38, 3193-3206.

Surry, P., Radcliffe, N., and Boyd, I. (1995). A multi-objective approach to constrained optimization of gas supply networks: The COMOGA method. In Fogarty (ed), *Evolutionary Computing*, Lecture Notes in Computer Science #993, Springer-Verlag, Sheffield, UK, 166-180.

Tamaki, H., Kita, H. and Kobayashi, S. (1996). Multi-objective optimization by genetic algorithms: A review. In Fukuda and Furuhashi (eds.), *Proceedings of the 1996 International Conference on Evolutionary Computation*, IEEE Press, New York, 517-522.

Tamaki, H., Mori, M., Araki, M., and Ogai, H. (1995). Multicriteria optimization by genetic algorithms: a case of scheduling in hot rolling process. In *Proceedings of the 3rd Conference of the Association od Asian-Pacific Operational Research Societies within IFORS*, World Scientific, Singapore, 374-381.

Tan, K., and Li, Y. (1997). *Multi-Objective Genetic Algorithm Based Time and Frequency Domain Design Unification of Linear Control Systems*. Technical Report CSC-97007, Department of Electronics and Electrical Engineering, University of Glasglow, Glasglow, Scotland.

Tarascio, V. (1968). *Pareto's Methodological Approach to Economics*. University of North Carolina Press, Chapel Hill, VA.

Thornton, W. (1971). Luminosity and color-rendering capability of white light. *Journal of the Optical Society of America*, 61(9), 1155-1163.

Thornton, W. (1992a). Toward a more accurate and extensible colorimetry. Part 1 – Introduction. The visual colorimeter-spectroradiometer. Experimental results. *Color Research and Application*, 17, 79-122.

Thornton, W. (1992b). Toward a more accurate and extensible colorimetry. Part II – Discussion. *Color Research and Application*, 17, 162-186.

Thornton, W. (1992c). Toward a more accurate and extensible colorimetry. Part III - Discussion. *Color Research and Application*, 17, 240-262.

Thornton, W. (1998). Toward a more accurate and extensible colorimetry. Part VI – Improved weighting functions. Preliminary results. *Color Research and Application*, 23, 226-235.

Todd, D. (1997). *Multiple Criteria Genetic Algorithms in Engineering Design and Operation*. PhD thesis, University of Newcastle, Newcastle-upon-Tyne, UK.

Tseng, C., and Lu, T. (1990). Minimax multiobjective optimization in structural design. *International Journal for Numerical Methods in Engineering*, 30, 1213-1228.

Turner, B., Ury, M., Leng, Y., and Love, W. (1997). Sulfur lamps - Progress in their development. *Journal of the Illuminating Engineering Society*, 26, 10-16.

van Trigt, C. (1999). Color rendering, a reassessment. *Color Research and Application*, 24, 197-206.

van Veldhuizen, D. and Lamont, G. (1998). *Multiobjective Evolutionary Algorithm Research: A History and Analysi*s. Technical Report TR-98-03, Department of Electrical and Computer Engineering, Graduate School of Engineering, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio.

Vos, J. (1978). Colorimetric and photometric properties of a 2° fundamental observer. *Color Research and Application*, 3, 125-128.

Vos, J., Estevez, O., and Walraven, P. (1990). Improved color fundamentals offer a new view on photometric additivity. *Vision Research*, 30, 937-943.

Walter, W. (1971). Optimum phosphor blends for fluorescent lamps. *Applied Optics*, 10, 1108-1113.

Walter, W. (1978). Optimum lamp spectra. *Journal of the Illuminating Engineering Society*. 7(1), 66-73.

Wienke, P., Lucasius, C., Ehrlich, M., and Kateman, G. (1993). Multicriteria target vector optimization of analytical procedures using a genetic algorithm, part 2: Polyoptimization of the photometric calibration graph of dry glucose sensors for quantitative clinical analysis. *Analytica Chimica Acta*, 271, 253-268.

Wienke, P., Lucasius, C., and Kateman, G. (1992). Multicriteria target optimization of analytical procedures using a genetic algorithm. *Analytica Chimica Acta*, 265, 211-225.

Wilson, P., and Macleod, M. (1993). Low implementation cost IIR digital filter design using genetic algorithms. In *Proceedings of the IEE/IEEE Workshop on Natural Algorithms in Signal Processing*, Chelmsford, UK, 1-8.

Winston, W. (1994). *Operations Research: Applications and Algorithms*. Duxbury Press, Belmont, California.

Wright, W. (1928-29). A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society, London*, 30, 141-164.

Wright, W. (1981a). 50 Years of the 1931 CIE standard observer for colorimetry. Inter-Society Color Council, *AIC Colour 81*, Paper A3.

Wright, W. (1981b). The historical and experimental background to the 1931 CIE system of colorimetry. In *Golden Jubilee of Colour in the CIE*, The Society of Dyers and Colourists, Bradford, UK.

Wyszecki, G. Stiles, W. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*. 2nd edition, John Wiley & Sons, New York

Zitzler, E. (1999). *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. PhD thesis, Swiss Federal Institute of Technology, Zurich, Switzerland.

Zitzler, E., Deb, K., Thiele, L., Coello, C., and Corne, D. (Eds.) (2001). *Evolutionary Multi-Criterion Optimization, First International Conference, EMO 2001 Zurich, Switzerland, March 2001, Proceedings*. Lecture Notes in Computer Science (LNCS) Vol. 1993. Springer-Verlag, Berlin.

Appendix A: Additional Figures

Several additional figures are presented for completeness in this appendix.

Figures 37-39 are plots of the distribution of chromaticity coordinate error ($[x, y]_{filtered} - [x, y]_{target}$) from 150 runs of the GA, for each $\overline{T}^c$ for the sulphur lamp, high pressure sodium, and metal halide spectra, respectively. Each point represents error from an individual run. The horizontal and vertical bell curves are normal probability density functions with height normalized and scaled to fit within the axes, and with the same mean and variance as the x and y chromaticity error, respectively. The dashed line represents a three standard deviation MacAdam ellipse (MacAdam 1942; MacAdam, 1943; Brown, et al., 1956) for the target chromaticity. A one standard deviation MacAdam ellipse is the boundary for just noticeable difference (JND) in chromaticity under ideal viewing conditions (adjacent fields, unlimited observation time, photopic conditions, foveal viewing); the JND will be larger for less than ideal viewing conditions (Narendran, et al., 2000; Brown, 1952, 1957; Brown, and MacAdam, 1949). A four standard deviation MacAdam ellipse is the ANSI standard tolerance for lamp color difference (ANSI, 1996), although major manufactures stay within a three standard deviation limit (OSI, 2002).

Figures 40-42 are plots of the 50th, 25th and 75th percentiles of transmittance across all filters at each $\overline{T}^c$ for the sulphur lamp, high pressure sodium, and metal halide spectra, respectively.

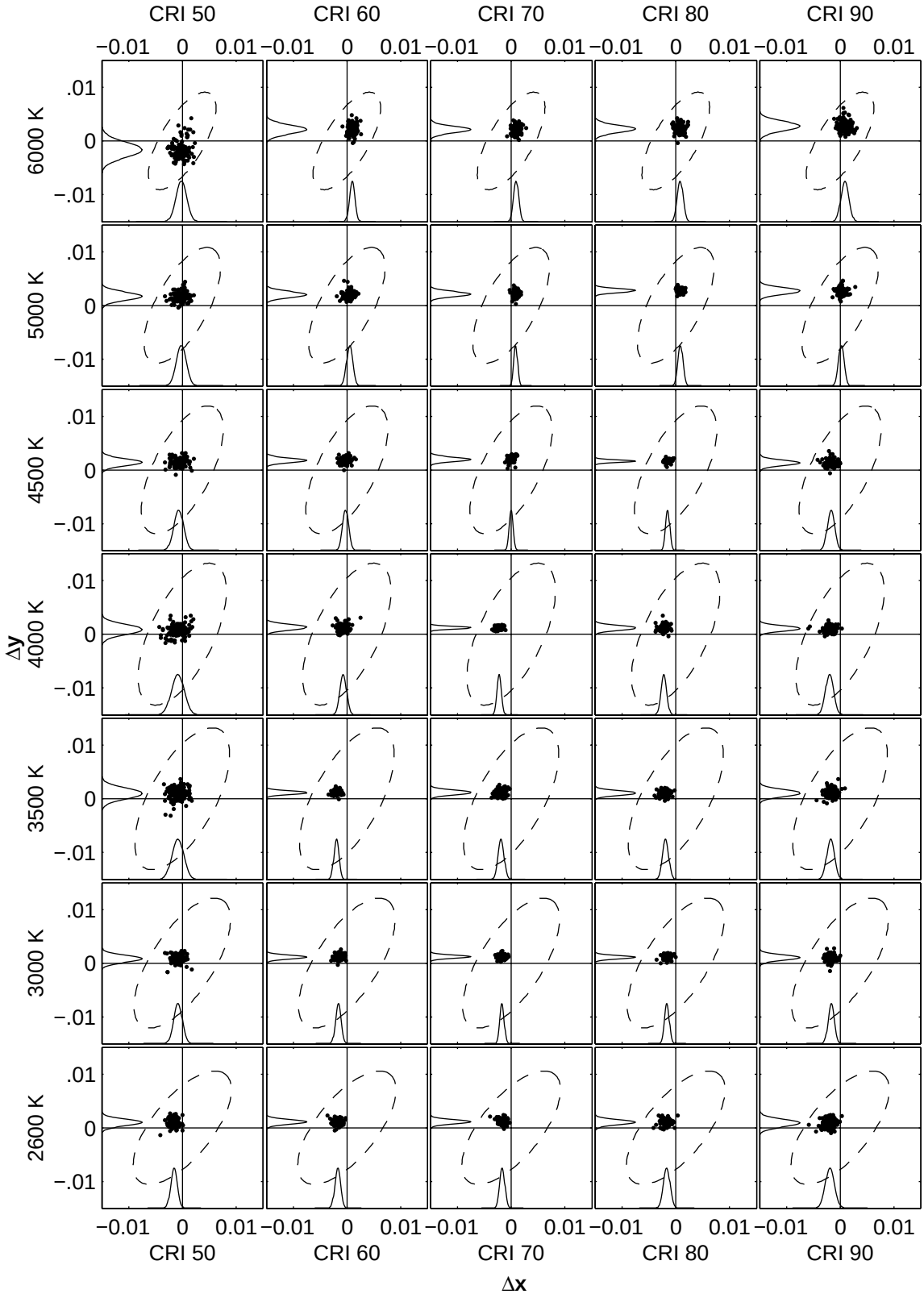

Figure 37. Distribution of chromaticity coordinate error by color temperature and CRI, sulphur lamp spectrum.

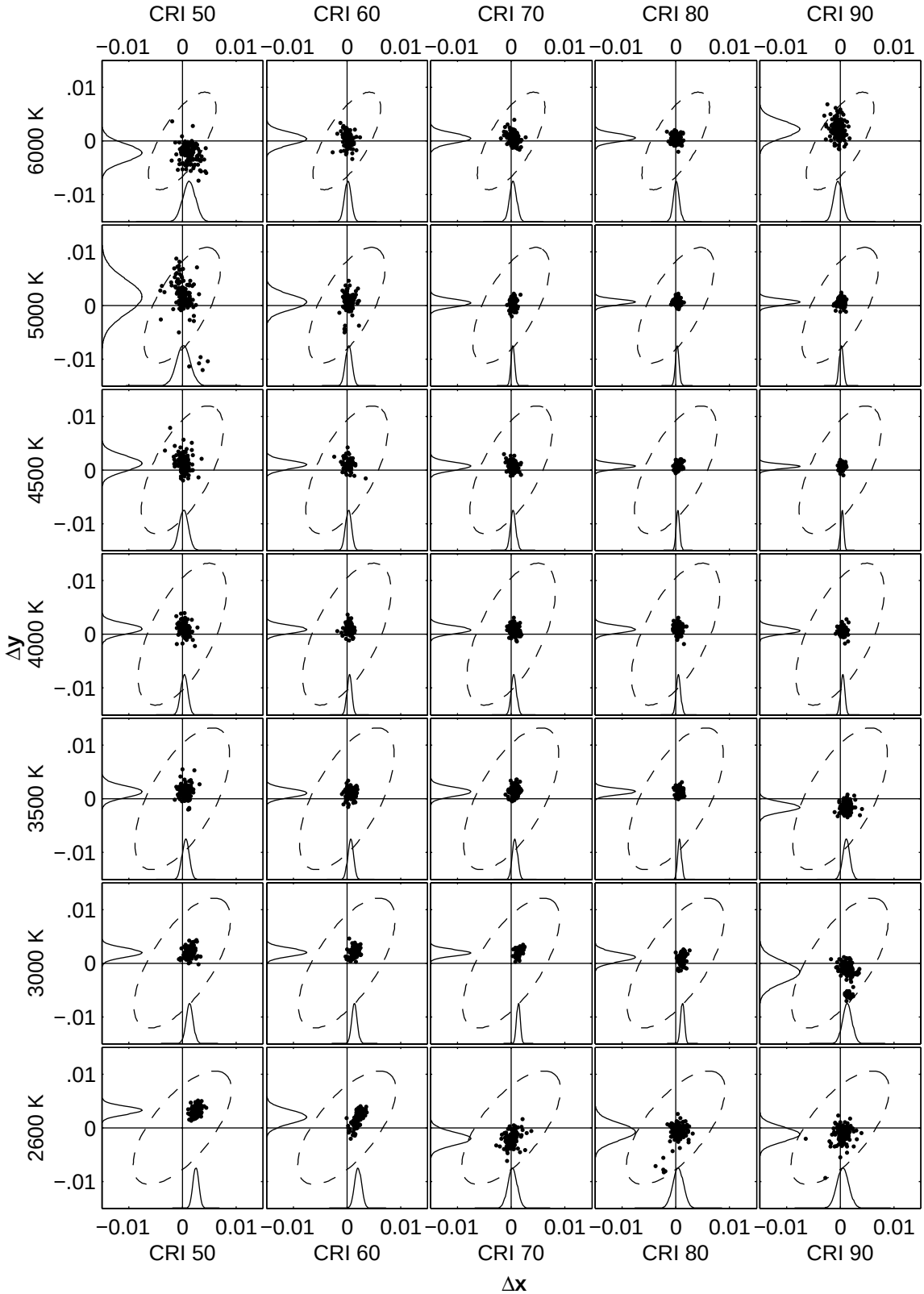

Figure 38. Distribution of chromaticity coordinate error by color temperature and CRI, HPS spectrum.

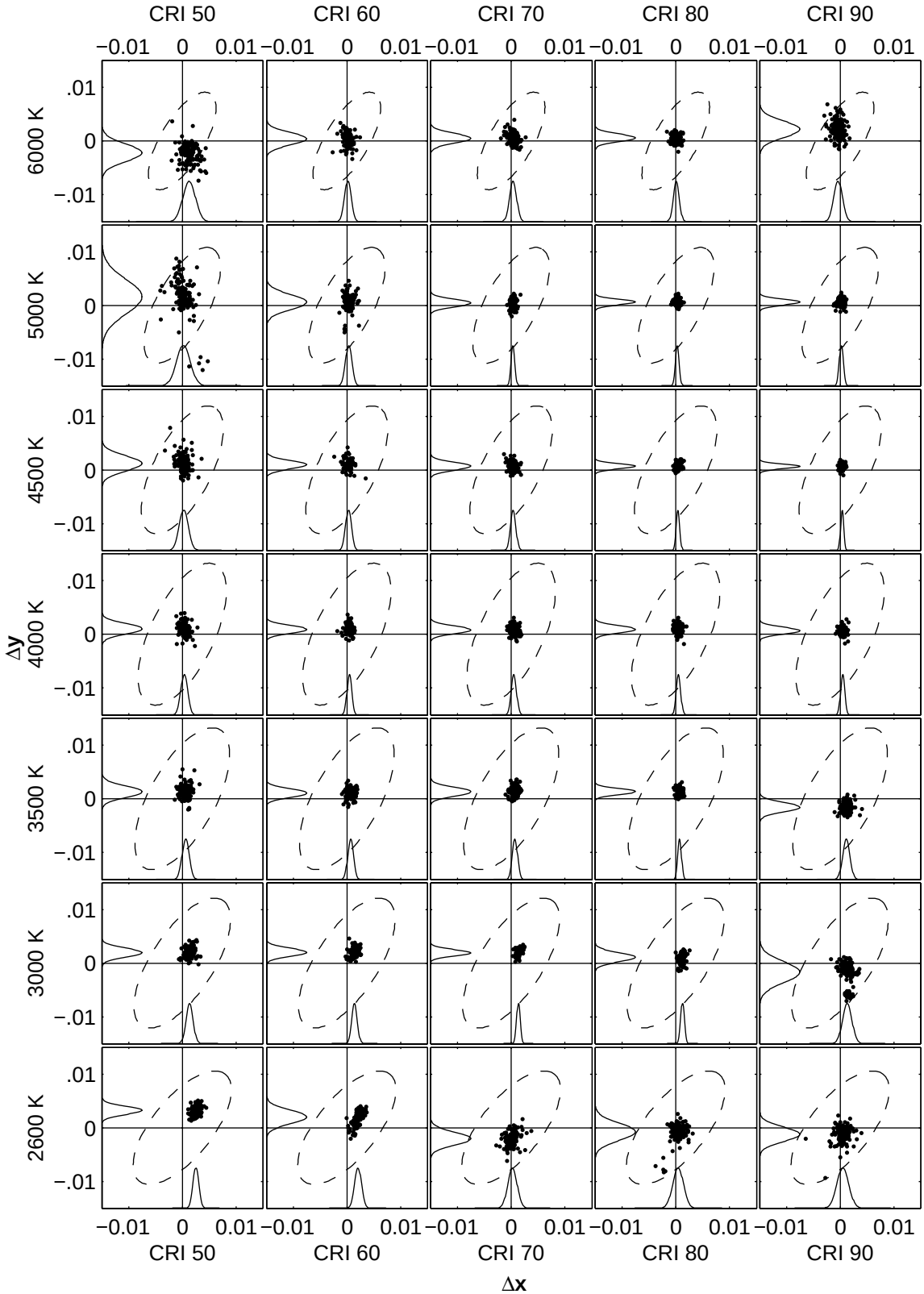

Figure 39. Distribution of chromaticity coordinate error by color temperature and CRI, MH spectrum.

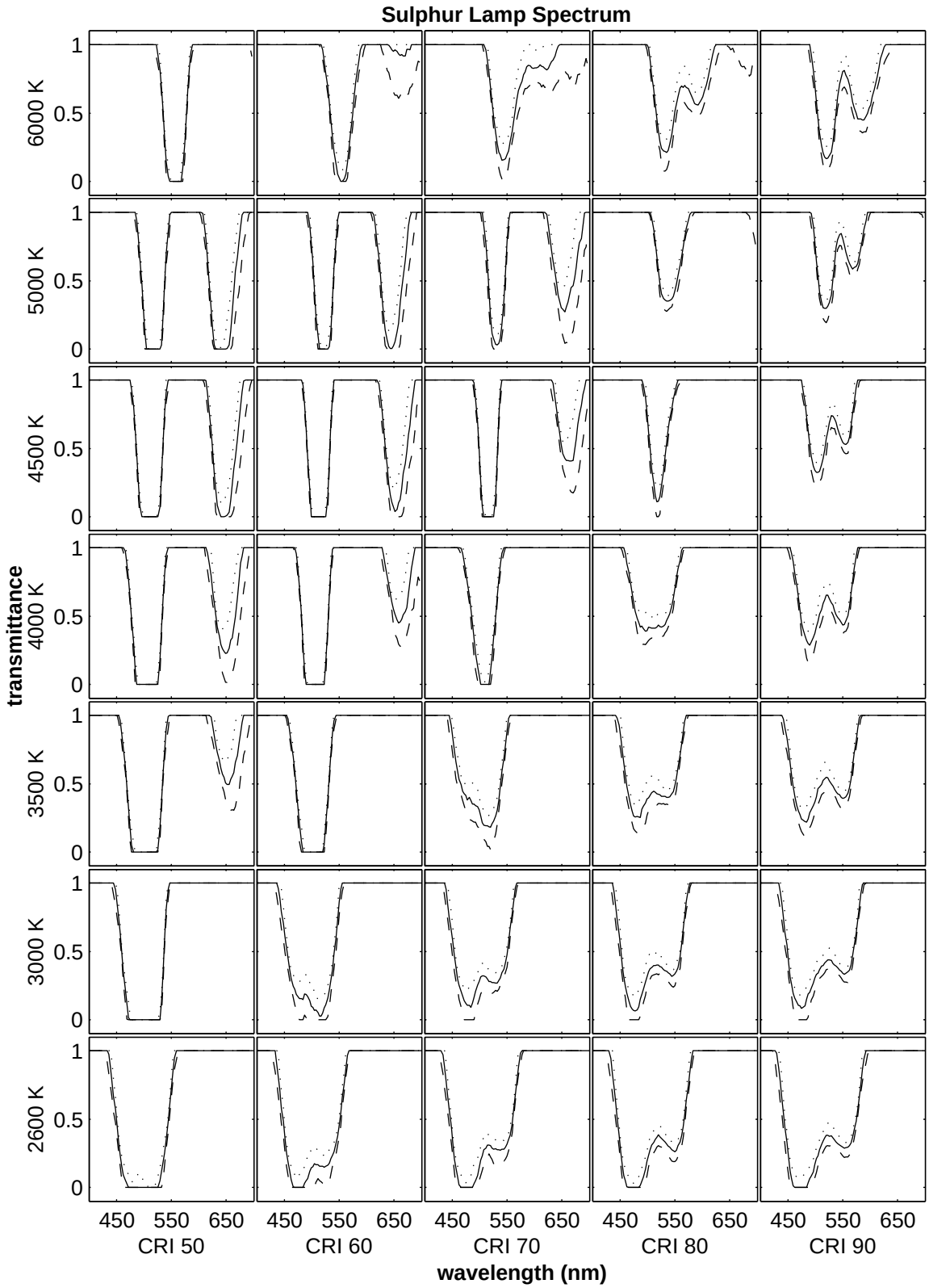

Figure 40. Median (solid line), 25th percentile (dashed line), and 75th percentile (dotted line) of the transmittance (at each wavelength sampled) of the filter designed for the sulphur lamp spectrum, at each $\overline{T}^c$.

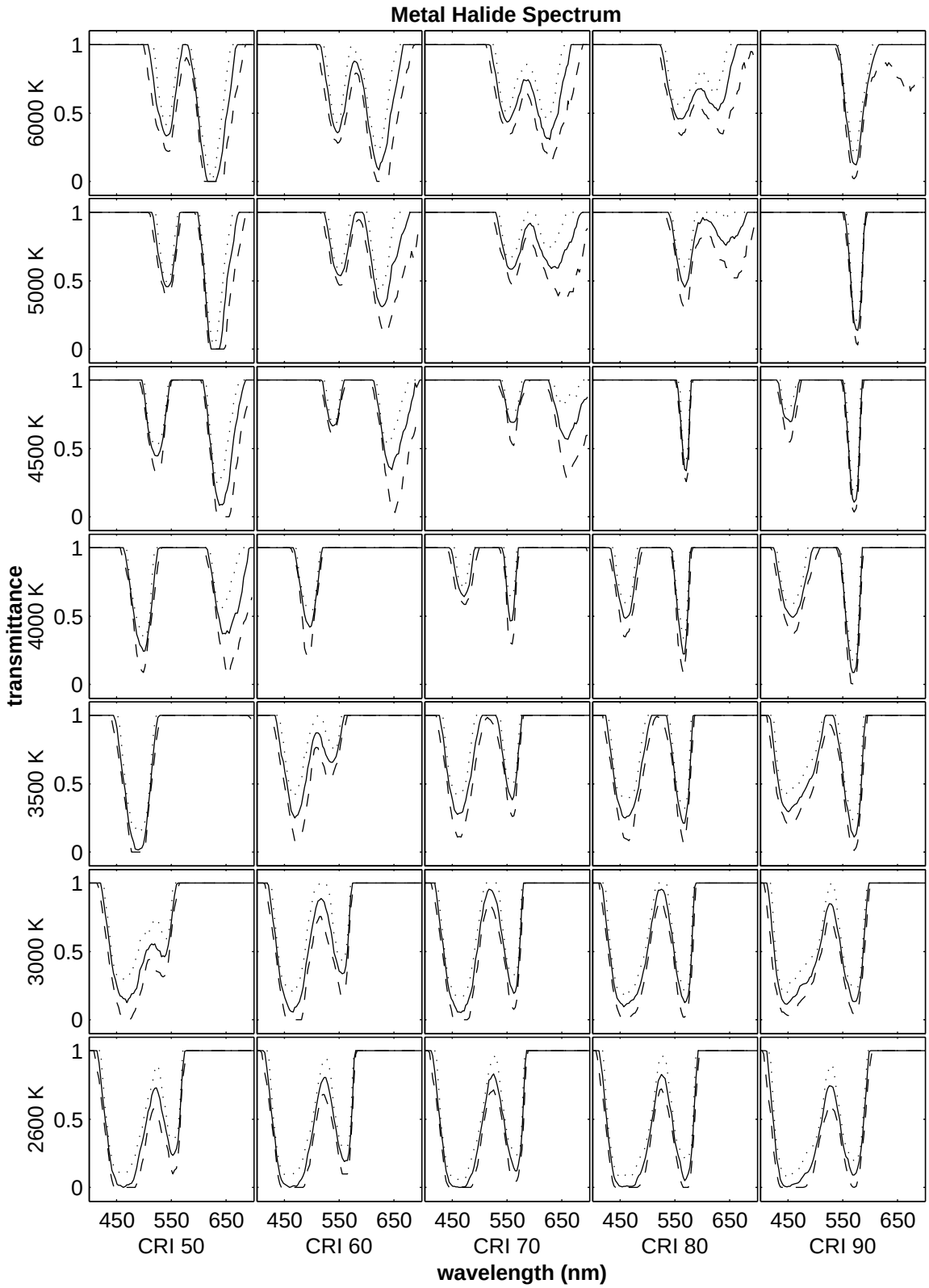

Figure 41. Median (solid line), 25th percentile (dashed line), and 75th percentile (dotted line) of the transmittance (at each wavelength sampled) of the filter designed for the MH spectrum, at each $\overline{T}^c$.

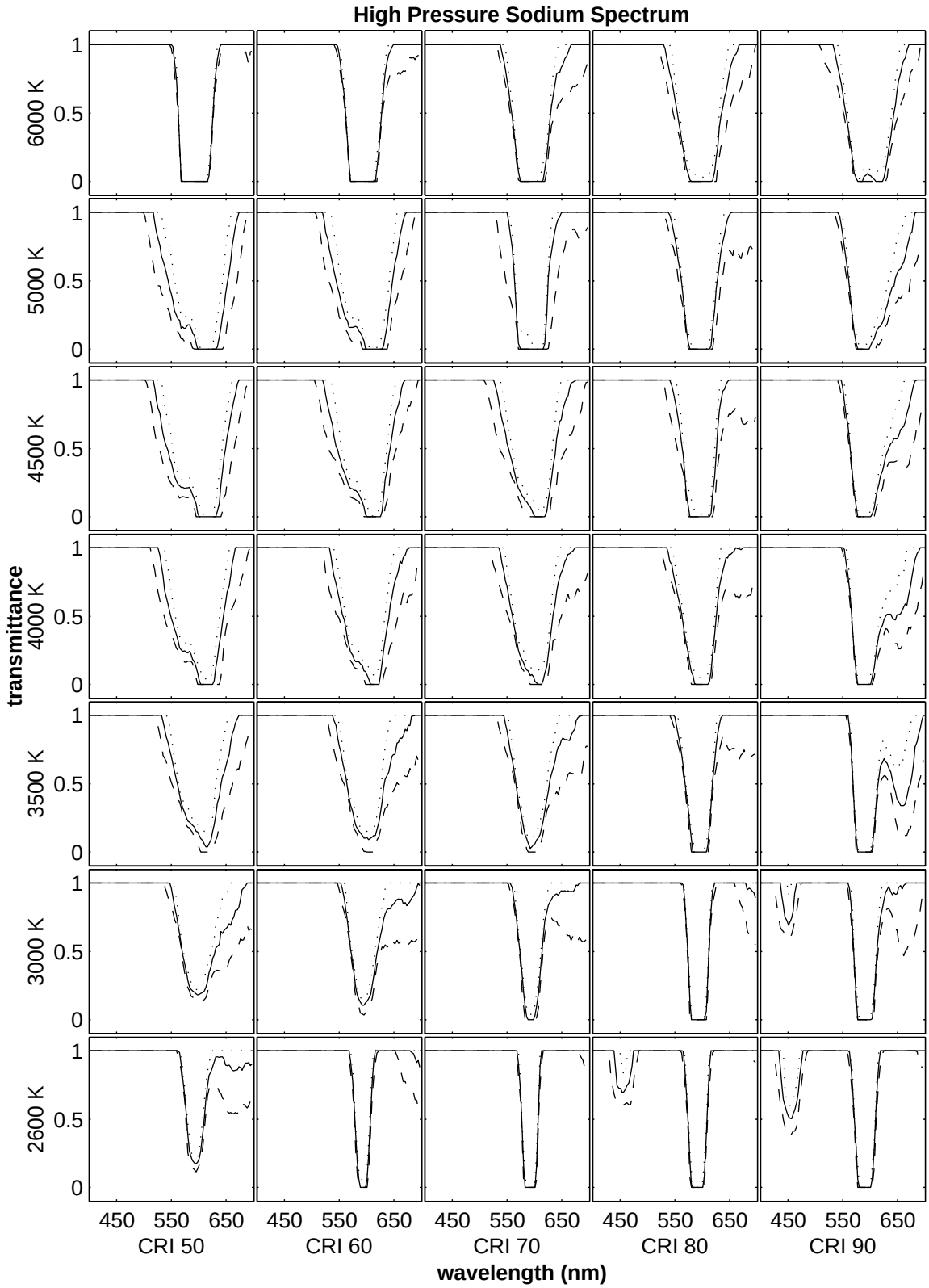

Figure 42. Median (solid line), 25th percentile (dashed line), and 75th percentile (dotted line) of the transmittance (at each wavelength sampled) of the filter designed for the HPS spectrum, at each $\overline{T}^c$.

Figures 43-45 are the same data shown in Figures 20-22, plotted in the perceptually uniform CIE 1976 (u',v') color space. Note that the difference between MacAdam's limit and the results generated using GAs does not appear to be an artifact of the perceptual nonuniformity of the 1931 CIE (x,y) chromaticity system.

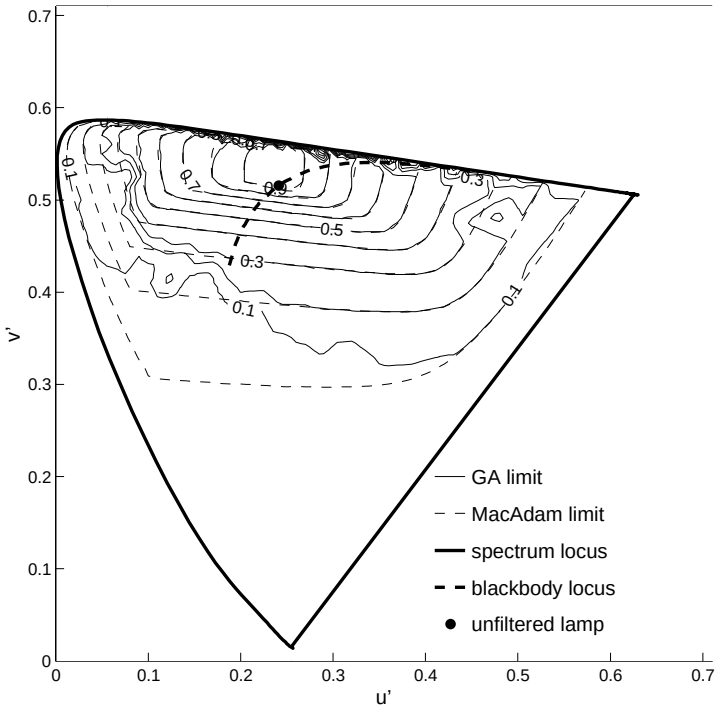

Figure 43. Efficiency contours for the incandescent source plotted in the CIE 1976 (u',v') coordinate system. Results from GA approach (solid line) are compared to the macadam limit (dashed line).

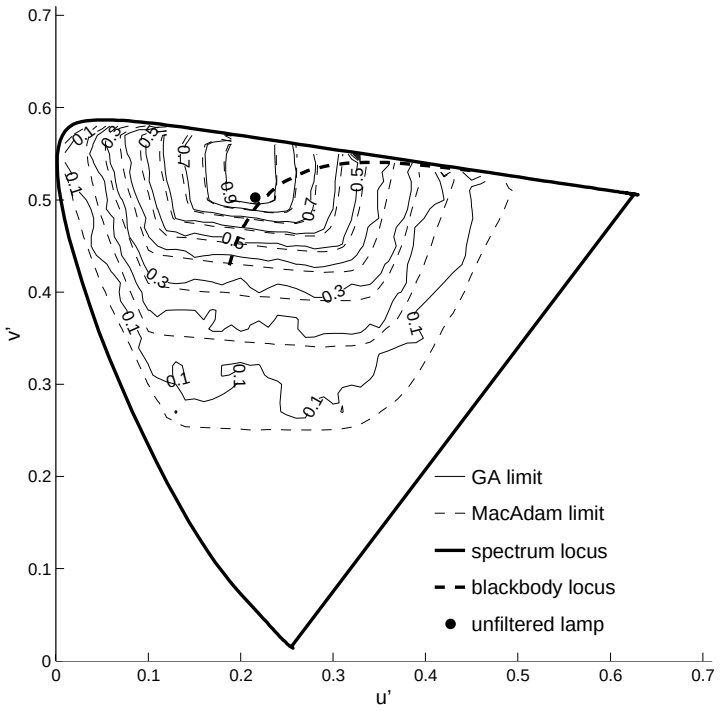

Figure 44. Efficiency contours for the MH source plotted in the CIE 1976 (u',v') coordinate system. Results from GA approach (solid line) are compared to the macadam limit (dashed line).

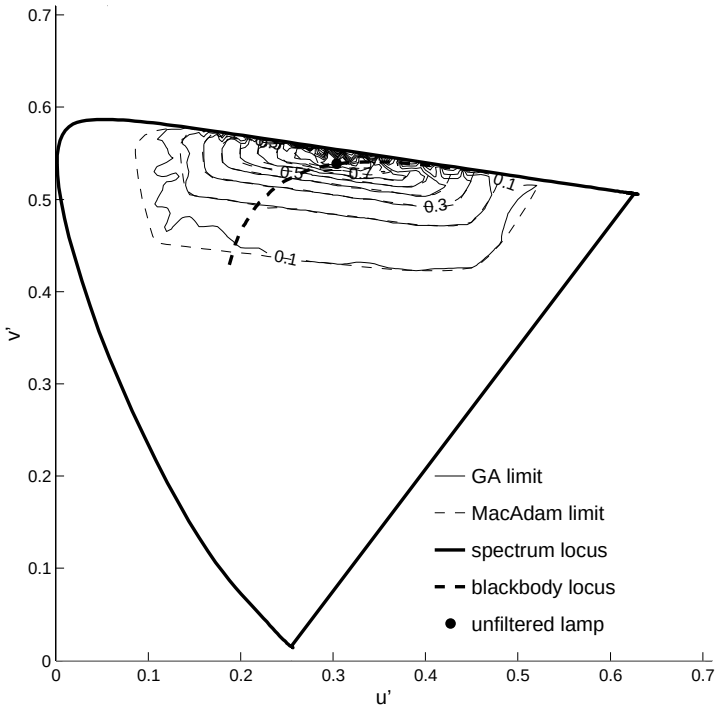

Figure 45. Efficiency contours for the HPS source plotted in the CIE 1976 (u',v') coordinate system. Results from GA approach (solid line) are compared to the macadam limit (dashed line).