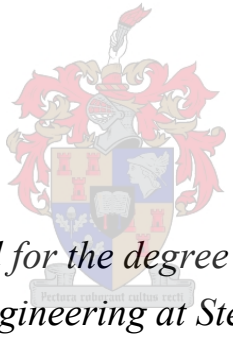
Geospatial Capacity Allocation Framework of Wind and Solar Renewable Generation for Optimal Grid Support

by

Chantelle Y. van Staden



Dissertation presented for the degree of Doctor of Philosophy in the Faculty of Engineering at Stellenbosch University.

Supervisor: Professor Hendrik J. Vermeulen

April 2022

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: April 2022

Abstract

Geospatial Capacity Allocation Framework of Wind and Solar Renewable Generation for Optimal Grid Support

C. Y. van Staden

Department of Electrical and Electronic Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Presented for the degree of Doctor of Philosophy in the Faculty of Engineering at Stellenbosch University.

April 2022

South Africa has displayed a unique energy supply profile over recent years, where the ability to consistently meet the energy demand has been constrained by physical limitations of the current energy supply infrastructure. The inadequate supply infrastructure results in countrywide loadshedding events, where total energy supply within high demand periods cannot be met. Low-grade coal, poorly maintained power plants and the impending decommissioning of existing thermal plants adds to the country's energy supply deficit. Inadequate supply in high demand periods typically requires response from expensive on-demand dispatch units, which are often non-renewable resources. This also equates to a decrease in grid supply stability. It is expected that optimised geospatial capacity allocation of new build wind and solar plants can assist in addressing the generation capacity constraints in the medium to longer term future. The framework proposed in this study favours a cascaded optimisation strategy, whereby the residual load profile is optimised statistically to reduce the requirements of ancillary services to complement baseload generation.

In support of a reliable future energy supply scenario with high penetration of renewable energy, the optimisation framework proposed in this work represents a probabilistic risk-based approach that seeks to minimise the number of events where high residual load values require ancillary service interventions to maintain power balance. In this approach, renewable energy resource features are categorised in terms of the statistical properties of the spatiotemporal wind and solar power profiles for a given set of daily and seasonal Time-of-Use periods. In this context, it is recognised that the resource characteristics and grid impact of wind and solar generation profiles can be interpreted with reference to the daily and seasonal cycles exhibited by the demand profiles, wherein some Time-of-Use periods are more important than others. Apart from the benefit of assigning renewable energy capacities to spatial regions rather than specific coordinates, clustering reduces the dimensions of input data sets dramatically. This reduces the dimensionality of the multi-variable optimisation search space, which translates to reduced risk of local minima and reduced computational cost.

The proposed framework has been implemented for a number of baseline case studies and optimisation case studies. It is concluded that the framework is highly flexible in the sense that the formulation of the minimum and maximum allocation constraints allow application for real-world scenarios where capacity allocation constraints apply on a regional level.

Overall, the optimisation framework provides a robust method for the geospatial capacity allocation of wind and solar resources. The framework employs a robust way of handling constraint scenarios when considering multiple highly granular resource clusters.

Uittreksel

Georuimtelike kapasiteitstoewysingsraamwerk van wind en sonkrag opwekking vir optimale netwerkondersteuning

("Geospatial Capacity Allocation Framework of Wind and Solar Renewable Generation for Optimal Grid Support")

C. Y. van Staden

Departement Eletriese en Elektroniese Ingenieurswese,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.

Proefskrif: Proefskrif ingelewer vir die graad Doktor in Filosofie in die Fakulteit Ingenieurswese aan die Universiteit Stellenbosch

April 2022

Suid-Afrika het die afgelope jare 'n unieke energie-voorsienings-profiel getoon, waar die vermoë om konsekwent aan die energievraag te voldoen deur fisiese beperkings van die huidige energie-voorsienings-infrastruktuur. Die onvoldoende voorsienings-infrastruktuur lei tot landswye beurtkrag-gebeurtenisse, waar die totale energie voorraad tydens hoë aanvraag periodes nie nagekom kan word nie. Laegraadse steenkool, swak onderhoud op kragsentrales en die naderende afskakel van bestaande termiese aanlegte dra by tot die land se tekort aan energie-voorsiening. 'n Onvoldoende aanbod tydens hoë-aanvraag-periodes vereis tipies 'n onmiddellike reaksie vanaf die kragopwekker, waar duurder intydse elektrisiteits-eenhede opgewek moet word. Hierdie eenhede is gewoonlik afkomstig vanaf nie-hernubare hulpbronne en plaas addisionele druk op krag-stelsel-stabiliteit.

Daar word verwag dat die beperkings op opwekkings-kapasiteit, in die medium- tot langtermyn toekoms, aangespreek kan word deur die geoptimaliseerde georuimtelike-kapasiteits-toewysing van nuwe wind- en sonkrag-aanlegte. Die raamwerk wat in hierdie studie voorgestel word, bevoordeel 'n kaskade-optimeringstrategie, waardeur die oorblywende-lasprofiel statisties geoptimaliseer word om die vereistes van bykomende dienste te verminder om basislading-opwekking aan te vul.

Ter ondersteuning van 'n betroubare toekomstige energie-voorsienings-scenario met 'n hoë penetrasie van hernubare energie, verteenwoordig die voorgestelde optimaliserings-raamwerk 'n risiko-gebaseerde waarskynlikheids-benadering wat poog om die aantal gebeurtenisse te minimaliseer waar hoë oorblywende laswaardes aanvullende diens-ingryping vereis om die kragbalans te handhaaf. In hierdie benadering word hernubare-energie-hulpbron-kenmerke gekategoriseer. Dit word gedoen volgens die statistiese eienskappe van die tydruimtelike wind- en sonkragprofiele, vir 'n gegewe stel daaglikse en seisoenale tyd-van-gebruik periodes. In hierdie konteks word erken dat die hulpbron-kenmerke van wind- en sonkragkragstelsels se opwekkings-profiele geïnterpreteer kan word met verwysing na die daaglikse en seisoenale siklusse, soos vertoon deur die aanvraag-profiel. In hierdie aanvraag-profiel is daar ook sommige tyd-van-gebruik periodes wat belangriker is as ander. Afgesien van die voordeel om hernubare-energie-vermoëns aan ruimtelike streke toe te ken, eerder as spesifieke koördinate, verminder die groepering van die insetdatastel-afmetings dramaties. Dit verminder die dimensionaliteit van die multi-veranderlike optimaliserings-soekruimte, wat neerkom op 'n verminderde risiko van plaaslike minima en berekenings-koste.

Die voorgestelde raamwerk is geïmplementeer vir 'n aantal basislyn-scenarios en optimaliserings-gevallestudies. Daar word tot die gevolgtrekking gekom dat die raamwerk hoogs buigsaam is rakende die

formulering van die minimum en maksimum toekennings-beperkings-toepassing, soos toegelaat vir werklike scenarios waar kapasiteits-toekennings-beperkings op 'n streeksvlak geld.

In die algemeen bied die optimaliseringsraamwerk 'n robuuste metode vir die georuimtelike-kapasiteits-toewysing van wind- en sonkragbronne. Die raamwerk gebruik 'n robuuste manier om beperkingscenarios te hanteer wanneer verskeie hoogs korrelvormige hulpbrongroeperings oorweeg word.

Acknowledgements

To God, my heavenly Father, my biggest motivator and source of unending love and support. Thank You for Your continuous blessings, for placing people in my life that I now call family and will love for eternity and finally, for helping me achieve more than I ever thought possible. What an amazing journey it has been, and I look forward to everything the future holds with You living in and through me.

You crown the year with Your goodness, And Your paths drip with abundance. - Psalm 65:11

In no particular order, I would like to express my sincere gratitude to the following people:

To my dear professor, Prof Vermeulen, words cannot express how grateful I am for your constant support and motivation. Thank you for always being there to help, mentor and inspire me. The depth of your knowledge and wisdom is unrivalled, and it has been an honour learning from you. Thank you so much, this would not have been possible without you.

Thank you to the support from my friends in E222, with the biggest thanks extended to Matthew Groch. Matty, thank you for all the great chats both in office and over skype while the pandemic had us cooped up indoors, for the Americano-with-cold-milk coffee breaks, for the overseas conference adventures and for all the help and motivation you have given me over the years, I am extremely grateful.

To Macyln Chingwena, for your incredible friendship and dissertation editing skills. Thank you for putting in the time to read and edit my dissertation, but most of all thank you for your support, love and motivation over the years.

To the three most amazing women I know, Juliette Janse van Vuuren, Estelle Pieters and Shannon Janse van Vuuren. You girls have been my rock, my home and my motivation. Mom, you are always loving, always kind and your support means more to me than you will ever know. Stellenbosch was a dream made reality because of you. To my Gran, Estelle, the amazing woman who introduced me to the world of Engineering. Thank you for believing in me from day one, for all of your love and for helping me, mom and Shan become the women that we are today. To my beautiful sister, Shan, thank you for all the laughs, the late-night shoulder massages while I am still sitting at my pc, for the cups of warm tea and the morning flap jacks. Being your sister is an honour and you have taught me so much through your kind and gentle nature. I love you all, to the moon and back.

To my dad, Wayne Janse van Vuuren, thank you for all of the love and support you have showed me throughout this PhD season. Your phone calls to check in and encourage me have made the journey that much more enjoyable and memorable.

To my new parents, Thys and Hannetjie van Staden, thank you for your support, acceptance and love. I appreciate all the encouragement and for understanding when I had to work on weekend visits.

To my GraceLife family, you have been a huge part of this success. I am so grateful that God gave me a family of believers who are full of love, kindness and grace.

To friends, who have become family! Sumie, Natasha, Terusa, Robyn, Jerusha, Karen, Armand and Lukas, thank you for all the coffees, snacks, cake breaks, flowers and constant words of encouragement and love. You are all are a major part of this journey and for that, I am so grateful.

To my dearest friend, Martinette Port, words cannot really comprehend the amount of love and support you have poured into me over this season. I will forever grateful be that you came into my life and made it 100 million times better. I have you to thank for my days starting with an encouraging message/voice note and my evenings ending with words of motivation and love. Thank you so very much.

I extend my gratitude to all my friends, family members, and university staff who I did not specifically mention by name. I sincerely appreciate all of your support.

Dedication

To my husband, Paul van Staden.

You are my biggest blessing, my best friend and my motivation to push to the end. Without your constant love and often pushing me into my office to work, this PhD would not have seen completion. Thank you for always showing me grace, kindness and for making me laugh even when it was a difficult part in the journey. I appreciate all the warm meals, shoulder massages, late night pep talks as well as all the dissertation editing. I am so grateful I got to marry my best friend in the middle of this PhD journey, and your surname deserves to be on the front page.

I love you and I dedicate this dissertation to you.

Table of Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

ANN	Artificial Neural Networks
CLARA	Clustering Large Applications
CM SAF	Climate Monitoring Satellite Application Facility
CSIR	Council for Scientific and Industrial Research
CSP	Concentrated Solar Power
DNI	Direct Normal Irradiance
DoE	Department of Energy
DSM	Demand Side Management
EE	Energy Efficiency
EIA	Environmental Impact Assessment
EM	Expectation Maximisation
ERA	Electricity Regulation Act
GA	Genetic Algorithm
GCCA	Generation Connection Capacity Assessment
IPP	Independent Power Producers
IRP	Integrated Resource Plan
LOLE	Loss of Load Expectation
LOLP	Loss of Load Probability
LPG	Liquefied Petroleum Gas
MCDF	Multi-Criteria Decision-Making Framework
MEDUSA	Model of Economic Dispatch and Unit commitment for System Analysis
NDP	National Development Plan
PAM	Partitioning Around Medoids
PV	Photovoltaic
PVGIS	Photovoltaic Geographic Information System
RE	Renewable Energy
REDZ	Renewable Energy Development Zone
REIPPP	Renewable Energy Independent Power Producer Procurement Programme
ROs	Research Objectives
SEA	Strategic Environmental Assessment
SOM	Self-Organising Maps

STC	Standard Test Conditions
ToU	Time-of-Use or Time of Use
WASA	Wind Atlas for South Africa

Chapter 1

Introduction

1.1 Research overview

South Africa is currently in the early stages of transitioning from non-renewable fossil fuels to Renewable Energy (RE) resources. During this transition period, grid penetration of utility scale wind and solar Photovoltaic (PV) resources are expected to increase substantially in the medium- and long-term future [1]. This places the South African energy market in a unique position, where optimal RE siting and grid integration planning can be methodically implemented and executed using best-practice approaches from countries with high RE penetration levels as a departure point [2, 3, 4].

It is broadly acknowledged that wind and solar resources are susceptible to a high degree of variability and sudden loss of generation capacity when unfavourable weather conditions are experienced [5, 6]. In instances where these unfavourable weather conditions affect a large area of highly localised RE resources, the System Operator may experience significant power-imbalance on the national energy supply gird. This is especially apparent in scenarios of high RE grid penetration percentages. Furthermore, the variability associated with RE generation gives rise to operational challenges in terms of reserve margins, scheduling and dispatch [7, 8].

Unlike the United States of America and Europe, which have large interconnected electricity grid systems, the South African electricity supply grid operates mainly as an isolated unit. The neighbouring countries of South Africa have limited generation capacity to support the South African electricity grid, thus, the energy trading market between the regional countries requires extensive further development. This emphasises the importance of medium- and long- term grid integration planning in the local context.

This research study focuses on the development, implementation and performance evaluation of a strategy for medium- to long-term optimisation of the capacity allocation mix and geographic spread of utility-scale wind and PV generation sources in the context of grid power balance. In this research study, the use of clustering techniques for the partitioning of temporospatial wind and PV generation profiles is introduced to aid in addressing plant siting issues concerning RE fleet management. An in-depth grid integration plan is also investigated through planning of the RE wind and solar resource fleet based on a probabilistic risk-reduced objective.

1.2 Research background and motivation

This section provides a detailed background of the current energy climate in South Africa. The country's current and future renewable energy policy and planning is also briefly discussed.

1.2.1 South Africa's renewable energy plan

The public nominal energy capacity for South Africa at the end of December 2020 was 51 600 MW, of which the nominal wind capacity was 2.5 GW, the solar nominal capacity was 2 GW and the combined nominal capacity of hydro, pumped storage and Concentrated Solar Power (CSP) summated to 3.8 GW [9]. This results in a RE grid integration ratio of 10.5 % RE capacity to 89.5 % thermal capacity, which still represents a relatively low penetration level of RE resources. In contrast, Denmark has a 32 % penetration of RE generation capacity, with an aim to increase the RE share to 43.6 % by 2021 [10].

Similarly, California has a 33 % RE generation capacity, with an ambitious goal of increasing RE generation to 100 % by 2045 [11, 12].

The South African National Development Plan (NDP) promotes future support and investment into South Africa's economic infrastructure. Since energy infrastructure is a vital component for a country's economic and social growth, the national grid network must be resilient enough to meet the anticipated future energy requirements of the industrial, commercial and private sectors. The NDP foresees that South Africa requires an additional 29 000 MW of generation capacity by 2030, of which 20 000 MW is to be supplied by renewable energy resources [13]. This capacity should be widely accessible, financially affordable and environmentally sustainable. The NDP formulated an Integrated Resource Plan (IRP), which was finalised and implemented in 2011, to support South Africa's future energy developments. However, it is recognised that the IRP requires continuous foresight and adaptability as the plan unfolds from the year 2010 to 2030. Many of the original assumptions have changed since the initial IRP was formulated. The demand projection did not increase as forecasted, and the performance of existing conventional thermal plants underachieved the availability factor initially assumed. The actual plant availability reduced substantially from the assumed 86 % to 71 % in the 2015/16 financial year, which recovered slightly to 77.3 % in the 2016/17 financial year [13]. This decrease in availability was the main contributing factor to the capacity limitations experienced in years 2011 until 2015. In the medium-term, it is critical that plant performance is enhanced and maintained to ensure supply security, since the Eskom fleet represents the main source of South Africa's generation supply capacity.

A revised and extended version of the original IRP was proposed in 2018, in which the planning period has been extended to 2050 [13]. The 2018 version specifies that 39 730 MW of new generation capacity must be developed, of which 18 000 MW has already been committed under the Renewable Energy Independent Power Producer Procurement (REIPPP) programme and the Eskom new build programme.

The 2019 IRP [14] projects a scenario where the generation capacity is expected to grow from 37 149 MW in 2018 to 77 834 MW by 2030. The energy mix for 2030 is expected to include 17 742 MW of wind energy and 8 288 MW of solar energy, i.e., 22.5 % and 10.5 %, respectively, as shown in Figure 1.1. The IRP projects an average annual demand growth of between 1.33 % and 2 % from a base of the order of 240 TWh from 2018 to 2030.

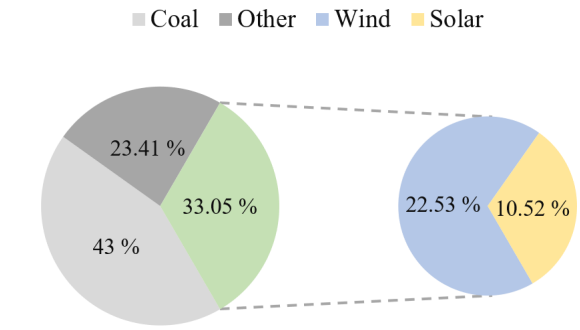


Figure 1.1: Generation fleet projected for the year 2030 [14].

The Department of Mineral Resources and Energy's Independent Power Producers (IPP) programme was established at the end of 2010 and is designed to support and facilitate investment from the private sector into RE generation through a competitive tender process. The introduction of privately owned, grid integrated RE plants is aimed at reducing governmental funding burdens, while introducing new generation technology, thereby meeting future energy demand with a green-energy approach. This process requires the bid winners to adhere to several stringent requirements. The bid submissions are

limited to onshore RE sources, namely, PV, wind, CSP, small hydro, biogas, biomass and landfill gas projects.

The South African energy sector is influenced by two acts, namely, the National Energy Act of 2008 (No. 34 of 2008) and the Electricity Regulation Act (ERA) of 2006 (No. 4 of 2006). Under the ERA, the Department of Energy (DoE) gazetted the electricity regulations on new generation capacity. These regulations set the guidelines and rules that are applicable when undertaking an IPP bid programme. The act also creates a fair, non-discriminatory environment between energy buyers and IPPs. The REIPPP allows the private sector to submit bids for renewable power plants and, if the tender is won, a 20-year power purchasing agreement is secured. The bids to produce privately owned RE plants are evaluated on a 70:30 ratio basis, where 70 % of the total bid value is allocated to the price per kWh of the energy produced and 30 % to the economic development criteria. This non-price criteria includes local job creation and community development as well as domestic industrialisation and black economic empowerment [15].

The first bid window, of seven planned bid windows, was initiated in August 2011, where 28 bidders of a total 53 were selected, and the round closed in November 2012. After the bid agreements were finalised in November 2012, the first projects were successfully online by November 2013. The REIPPP signed another 27 power purchase agreements in June 2018, which entails the addition of 19 400 MWs of RE (14.4 GW of wind and 6.0 GW of PV) by 2030 [16]. To date, 6 422 MW of electricity, from 112 IPP projects in seven bid rounds, has been procured. 5 078 MW of electricity, from 79 IPP projects, has been successfully connected to the national grid and is currently operational [17]. According to the latest IPP quarterly report, 37 696 MW of new and committed capacity is to be added between 2019 and 2030 [17]. Table 1.1 gives the total allocated capacity per bid window for each technology type. Figure 1.2 depicts the geographic locations of the completed plants for each bid window. Figure 1.3 depicts the locations of the REIPPP programmes wind and PV projects up to Bid Window 4B.

The siting considerations pertaining to the wind and PV generation fleet are typically addressed by the IPPs and the utility during the scope and design stages of individual plants. These considerations are predominantly determined by the costs associated with the various technologies, regional economic development objectives, availability and cost of land, existing transport infrastructure, grid connectivity, transmissions grid capacities, the location of load centres, etc.

Currently, the South African energy mix is failing to meet the supply and demand scenarios, equating to a lag in technological energy advancements in energy efficiency and power system reliably. Neglecting to adhere to the two-yearly revision called for by the IRP, the DoE has frustrated the energy sector in the planning and adaption of the future energy vision as the country expands. However, with the updated IRP 2019, a revised roadmap for generation capacity and RE integration in South Africa, marks a promising step in the planning and implementation process.

The latest version update of the South African IRP was gazetted in October 2019, where the countries energy integration planning and forecast has been adjusted and updated [14]. The IRP proposes planning updates to the electricity generation fleet, to be implemented between 2020 and 2030. The IRP addresses how the countries energy demand should be met in terms of generation technology, capacity, cost and projected implementation time. The IRP research is grounded on a Multi-Criteria Decision-Making Framework (MCDF), which modelled various outcomes based on multiple electricity system scenarios. The scenarios include factors such carbon dioxide emission limits and the associated carbon taxes, generating build delays, regional development of diverse electricity import options as well as improved demand side management. Within the various scenarios, the MCDF includes parameters such as the cost of energy, the actual energy emissions, natural resource consumption, etc. The resulting plan represents a balance between cost, risk and key constraints.

Table 1.2 displays the projected energy plan for South Africa up until 2030. By 2030 the planned renewable energy penetration percentage from wind and solar resources alone is 33.50 %. This is a significant penetration increase, which requires further intricate reliability and logistical planning.

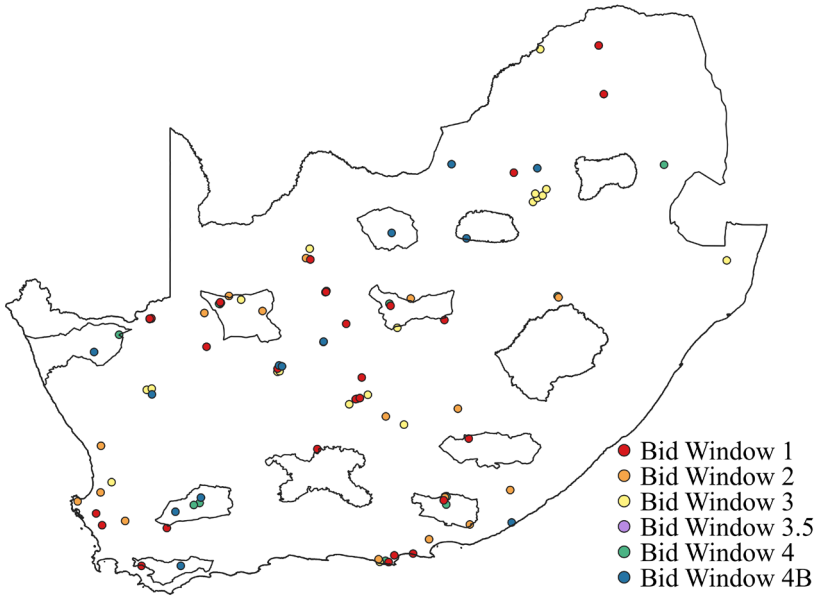


Figure 1.2: Locations of the allocated renewable energy plants for each bid window.

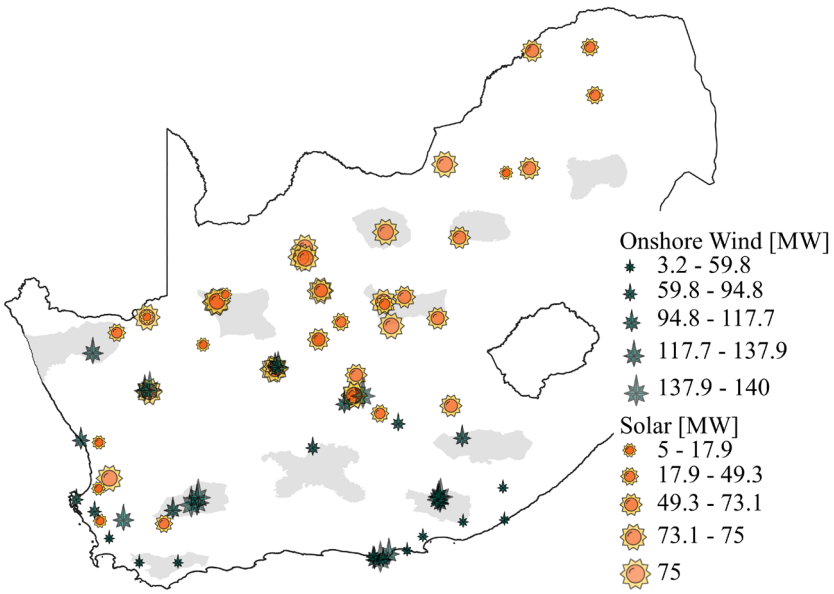


Figure 1.3: REIPPPP wind and PV projects.

Table 1.1: The REIPPPP capacity allocation by technology and bid window [18].

		Wind	PV	Solar CSP	Landfill Gas	Biomass	Small Hydro
BW 1	Capacity MW	649	627	150			
	No. of projects	8	18	2			
BW 2	Capacity MW	559	417	50			14
	No. of projects	7	9	1			2

BW 3	Capacity MW	787	435	200	18	17	
	No. of projects	7	6	2	1	1	
BW 3.5	Capacity MW			200			
	No. of projects			2			
BW 4	Capacity MW	676	415			25	5
	No. of projects	5	6			1	1
BW 4B	Capacity MW	686	398				
	No. of projects	7	6				
Total	Capacity MW	3357	2292	600	18	42	19
	No. of projects	34	45	7	1	2	3

Table 1.2: The updated long-term Integrated Resource Plan 2019 [14].

	Coal		Nuclear	Hydro	Storage	PV		Wind	CSP	Gas & Diesel	Other
Current base	37149		1860	2100	2912	1474		1980	300	3830	499
2019	2155	-2373						244	300		
2020	1433	-557				114		300			
2021	1433	-1403				300		818			
2022	711	-844			513	400	1000	1600			
2023	750	-555				1000		1600			500
2024			1860					1600		1000	500
2025						1000		1600			500
2026		-1219						1600			500
2027	750	-847						1600		2000	500
2028		-475				1000		1600			500
2029		-1694				1000		1600			500
2030		-1050				1000		1600			500
TIC	33364		1860	4600	5000	8288		17742	600	6380	
TIC (%)	43		2.36	5.84	6.35	10.52		22.53	0.76	8.10	
AEC (% of MWh)	58.80		4.50	8.40	1.2*	6.30		17.80	0.60	1.30	

	Installed Capacity
	Committed/Already Contracted Capacity
	Capacity Decommissioned
	New Additional Capacity
	Extension of Koeberg Plant Design Life
	Allocation to the extent of the short-term capacity and energy gap. Includes Distributed Generation Capacity for own use
TIC	Total Installed Capacity
AEC	Annual Energy Contribution

1.2.2 The South African Renewable Energy Development Zones

In a recent development, a Strategic Environmental Assessment (SEA) study carried out by the Council for Scientific and Industrial Research (CSIR) of South Africa identified eleven Renewable Energy Development Zones (REDZ), shown in Figure 1.4. The REDZs were identified for the efficient deployment of future utility-scale wind and PV plants [1]. Initially, 8 REDZs were gazetted on the 16th of February 2018 and as of the 26th of February 2021, the Department of Environment, Forestry and Fisheries has gazetted three additional geographic areas for the planning of large-scale wind and PV facilities [19].

These geographic regions are dispersed throughout South Africa and were determined with reference to various factors, including RE yield potential, restricted heritage sites, socioeconomic circumstances, agriculture terrain, landscape and biodiversity. The proposed zones require a less extensive Environmental Impact Assessment (EIA) for new builds, therefore decreasing the EIA assessment time for wind and PV plant placement from 300 days to 147 days. This incentives IPPs to site RE plants within these zones. Although the introduction of the REDZs is expected to introduce a degree of geographic dispersion in the future siting of RE plants, the current scenario still lacks a coherent approach towards optimising the RE capacity allocation mix and the locations of future plants in the context of national grid support and power balance. The eleven REDZs are strategically placed within 5 powerline corridors. These energy corridors were gazetted as powerline corridors in February 2018, after also being identified in the electricity grid infrastructure SEA study completed in 2016. Figure 1.4 depicts the eleven REDZs. Figure 1.5 shows the 5 energy corridors, together with the existing and planned powerlines.

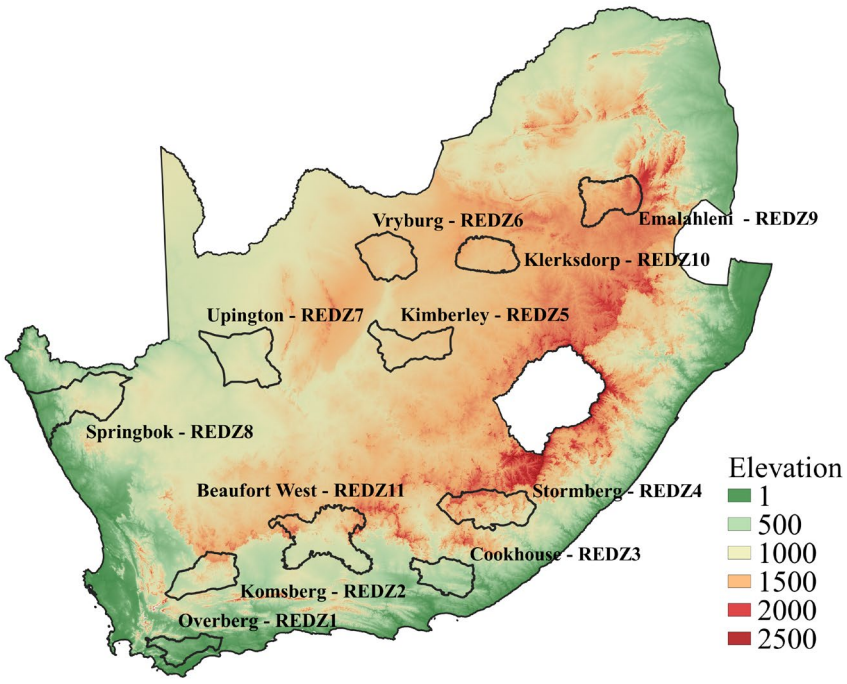


Figure 1.4: The eleven South African Renewable Energy Development Zones.

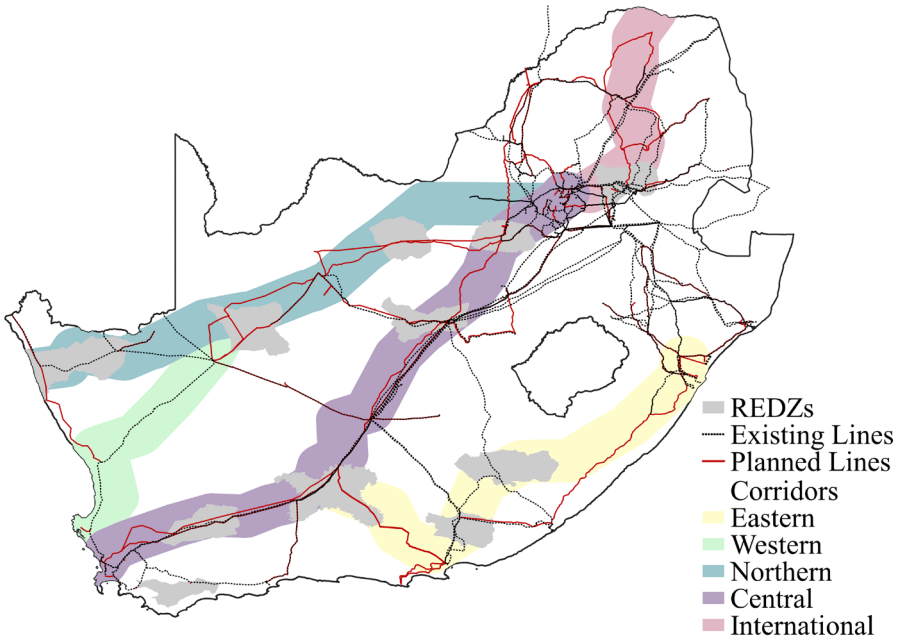


Figure 1.5: The eleven Renewable Energy Development Zones and the five energy corridors, with the existing and planned powerlines.

1.2.3 Grid impacts of renewable energy generation

In this section, the key performance indicators related to increased renewable energy penetration are expanded on, including power system security, power system adequacy and capacity credit of RE power plants.

The integration of RE resources onto the electricity grid has the potential to provide environmentally friendly and cost-effective solutions to the increasing energy demand. The introduction of these low-emission resources has led to many countries setting ambitious RE penetration targets for the medium- and long- term future. However, the temporal power generation profiles of wind and PV resources exhibit a highly stochastic profile, which could translate to adverse impacts on grid operations. These impacts are specifically noticeable in operational aspects such as capacity availability, system stability, ramp rates, forecasting, optimal dispatch, scheduling and power quality, including power system transients and harmonic distortion [20, 21]. It follows that the variable nature of RE resources require changes to power system planning and operations to successfully meet the targeted integration demands.

The effective capacity value of wind or PV generation depends on the integration and planning processes that support and maximise the availability of reliable capacity from these RE resources. Optimised capacity value of wind and solar energy is achieved when the use of conventional fossil-fuel-based generation is decreased and replaced by RE sources during high demand periods. Power system planning should ensure that the RE resources provide adequate generation capacity to meet the demand.

Should accurate planning and demand forecasting be neglected, the power system runs the risk of having insufficient capacity to meet the demand. This scenario typically requires response from expensive on-demand dispatch units, which are often non-renewable resources. This also equates to a decrease in grid supply stability. However, the capacity value of RE generation can be maximised by aligning the generation profile with the demand profile [22].

Wind energy compared to PV energy, exhibits a resource profile with greater variability. The wind energy resource is probabilistically just as likely to display high power outputs in peak demand periods

as it would in off-peak periods. For this reason, wind and PV resources should be aggregated as an interconnected system and allocated to geographically dispersed areas. This mitigates the risk associated with sudden changes in weather patterns, while introducing counteracting variability traits to produce a more reliable capacity output.

If the RE penetration level increases substantially, without stability planning, and the demand becomes reliant on the stochastic RE supply, where a sudden loss in capacity may result in a blackout condition [23].

An increase in grid connected stochastic RE generation induces changes in the demand profile. This can increase the frequency and severity of the subsequent ramping effects. This is especially noticeable with increased PV resource integration, where the net demand is reduced significantly during the midday period. This results in significant loss of capacity outside of the daily solar irradiation cycle. This effectively accentuates the peaking effects of the demand profile in the early morning and evening periods. The effect induced by this phenomenon is illustrated by the *Duck Curve* effect in Figure 1.6, which graphically depicts the effects of increasing predominately PV penetration capacity on the demand profile [3, 24].

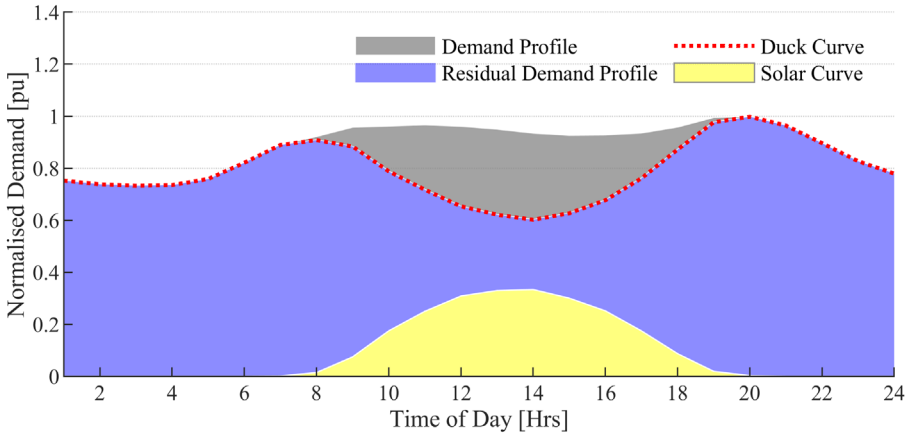


Figure 1.6: The *Duck Curve* effect on the residual demand, with high PV penetration capacity.

Economic dispatch is also an important aspect to consider for optimal power system operations. Economic dispatch involves dispatching generation resources such that the cost of generation is minimised. In this context, the integration of renewable energy resources increases the need for advanced forecasting and faster scheduling technologies to assist in mitigating the impacts of the stochastic nature of RE on network stability, network security and the overall generation profile. Implementation of an effective demand management system can assist in achieving optimal economic dispatch, which supports energy balance on the power system and allows for an optimal mix of energy generation while minimising operational costs [25]. Countries such as California and Spain, which exhibit high levels of RE capacity, often experience an undersupply or oversupply in certain demand periods if not managed correctly [2, 3, 4]. Optimised demand dispatch allows for the minimisation of wasted green-energy due to otherwise necessary energy curtailment methods [26].

A common consequence of the intermittent nature of wind and PV resources is voltage fluctuation [27]. The fluctuation in voltage levels leads to a phenomenon called flicker [28], which affects sensitive electronic and electrical equipment on weak grid systems and decreases equipment lifespan [29]. Another RE integration issue seen by the System Operator is harmonic distortion. Harmonic distortion occurs when a nonlinear load draws non-sinusoidal current waveforms, e.g., pulsed current waveforms with high amplitudes, from the supply voltage, creating a condition where current is disproportional to the

voltage supply. This creates distortion of the supply voltage waveform [30]. If the injection of the operating harmonics is not smoothed within the accepted tolerance, harmonic distortion occurs [31].

1.2.3.1 Power system reliability

Power system reliability is the overall ability to meet the demand regardless of plant failures or unexpected supply reduction occurrences. Billinton and Allan [32] defines the reliability of the electric power system as the ability to economically satisfy the system load with reasonable assurance of quality and continuity. The efficiency and ability to which the electric power system fulfils this responsibility can be conceptualised as the power system reliability. The concept of power system reliability can be described and subdivided into two aspects, namely, system adequacy and system security.

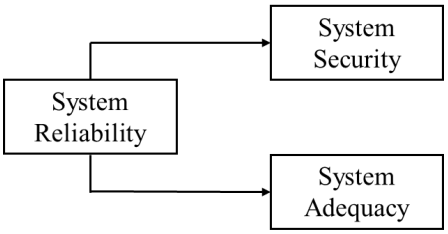


Figure 1.7: Power system reliability subdivisions.

System security pertains to the dynamic and transient system disturbances which occur within short time frames, from minutes to an hour. System adequacy describes a power system with the existence of sufficient facilities to meet demand requirements within the steady state limits. The system adequacy defines the power systems static conditions and is pertinent to the long-term system planning time frame and does not include system disturbances.

Since system reliability covers a large and complex range of highly integrated networks, the power system itself is divided into three system sublevels. These system sublevels include power generation, transmission, and distribution. Focusing on the long-term power system planning and system adequacy, the assessment of the system sublevels is conceptualised into hierarchical levels.

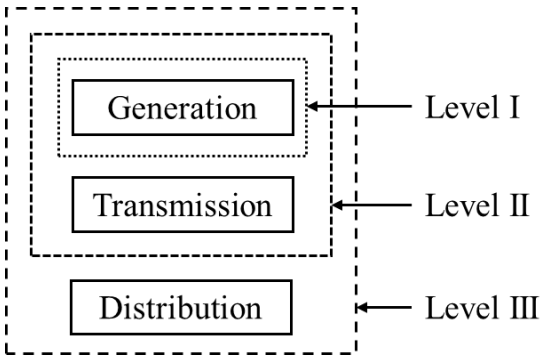


Figure 1.8: Hierarchical levels of power system adequacy [32].

The hierarchical level I consists solely of the generation aspect, level II includes both the transmission and the generation aspects and level III encompasses all the system sublevels [32].

The adequacy evaluation at hierarchical level I is described as the ability to meet the aggregated system load requirements and is referred to as the generating capacity adequacy [33]. This limits the adequacy evaluation for this hierarchal level to the estimation of necessary generating capacity to meet the system demand and to ensure adequate capacity to complete preventive and corrective maintenance on the generation facilities.

The Loss of Load Probability (LOLP) is the probability of not meeting the full load demand at a certain point in time and can be mathematically described as

$$LOLP_i = P(C_i < L_i), \quad 1.1$$

where $P(C_i, L_i)$ is the probability of the available capacity, C_i, being less than the daily peak demand, L_i, at time i. The Loss of Load Expectation (LOLE) is the number of hours the load will not be met in a specified time frame. The LOLE can be mathematically equated to the summation of LOLP,

$$LOLE = \sum_{i=1}^{n} LOLP_i, \quad 1.2$$

for n days of the year.

With the introduction of variable RE generation technology, the generic way of calculating system adequacy must be adapted to account for power delivery which relies largely on weather patterns rather than mechanical availability. The mechanically driven power generation systems are traditionally modelled as a binary distribution, where the generation is either available or unavailable.

Ibanez and Milligan [34] addresses the need for quantifying the capacity credit or capacity value of an inherently stochastic resource. They deliberated that the loss of load in the LOLE metric may display some inaccuracies for the modern interconnected systems. Three methods for calculating the capacity value for wind, PV and CSP are explored. The capacity value is defined as the effective load carrying capability of each generator type for each reliability metric, namely the *loss of load expectation*, *expected unserved energy* and *loss of load hours*. The study concluded that the three probability-based metrics compared show high correlation between results all round.

1.2.3.2 Renewable energy capacity credit

Capacity credit can be described as the contribution quantification of a generation technology to support the energy demand, otherwise defined as the generation adequacy of a power system. The concept of capacity credit has gained importance in recent years, with conventional and renewable energy technologies having vastly different demand support capacities. There are various methods that can be used to define and calculate the capacity value for a specific generation technology. The use of capacity value calculations can give insight into an effective margin of capacity necessary within a system which has a high percentage of RE penetration. Comparing the capacity credit of a conventional power plant and a wind power plant, assuming an equal annual power output across generation types, a wind plant has a lower contribution to the power system generation adequacy compared to a conventional power plant. The reliability of a wind plant supply would therefore need to be supplemented by back-up units [35].

There are multiple ways of calculating the capacity credit of a generating technology, Amelin [36] compares four such topical methods. These methods include the equivalent firm capacity method, the guaranteed capacity method, the equivalent conventional power plant method, and the effective load carrying method.

1.2.4 Impacts of renewable energy flat feed-in tariff structures

A flat feed-in tariff structure currently applies in South Africa for utility-scale RE generation. The focus of optimising the financial viability of a RE plant thus encourages a scenario where the siting of RE plants by IPPPs is based predominantly on achieving the highest cumulative energy yield and optimising the capital costs associated the grid connection, technology and construction costs [15]. In considering the fleet of utility-scale RE plants currently associated with the REIPPP program, the present scenario has given rise to a relatively high degree of geographic concentration of RE plants. This is predominately a result of the prevailing flat feed-in tariff structure. However, the flat feed-in tariff does not take cognisance of the true generation costs associated with the diurnal and seasonal Time-of-Use (ToU)

periods. The local utility implements a diverse range of ToU tariff structures for the various consumer sectors, which embody the cost of generation in the context of the national demand profile. Table 1.3 shows the ToU periods and associated energy costs for the high demand season ranging from the 1st of June to 31st of August, and the low demand season ranging from the 1st of September to 31st of May, of the Megaflex Generation tariff structure [37]. These costs exhibit a high degree of variability across the daily ToU periods and the two demand seasons. This suggests that there is a business case for financial incentives to be offered to IPPs to locate RE plants such that the residual load during expensive ToU periods is reduced, rather than maximising the cumulative annual energy yield, as inspired by a flat feed-in tariff.

Table 1.3: Normalised cost of energy for the Megaflex ToU tariff system [37].

Demand period	Low Demand Season	High Demand Season
Off-peak	0.14	0.17
Standard	0.23	0.30
Peak	0.33	1

1.3 Defining the research

South Africa occupies a large geographic area, with highly suited terrain and climatic conditions for the large-scale integration of RE resources [38]. The introduction of a location specific integration plan, based on these resources is of paramount importance [39].

1.3.1 Research question

The main research question pertaining to this research project can be formulated as follows:

> *Can a comprehensive RE geospatial capacity allocation framework be developed that incorporates a risk-based approach to grid support, taking cognisance of real-world regional siting constraints, using geospatial wind and solar resource clusters as the input dataset?*

1.3.2 Research aim

Wind and PV resources have become the leading green technologies of the RE sector, where the need for transitioning to a more sustainable energy supply has resulted in a surge in RE investment interest. However, from a grid perspective these energy sources are non-dispatchable and exhibit a high degree of variability and uncertainty. The variability associated with the combined wind and PV generation fleet can be substantially mitigated by optimising spatially [40, 41]. The research aim is to develop a RE resource planning framework to be used by government, public electric utility as well as private electric utility. The framework should consider the current state of the energy generation fleet, the projected energy integration capacities as well as the current abilities and limitations of the supply infrastructure to increase reliability in the future energy supply fleet. The framework should produce clustered regions which define the underlying resource in view of optimised grid supply objectives as well as spatially optimised RE capacity allocation maps.

1.3.3 Research objectives

This research targets the development, implementation and performance evaluation of a RE planning strategy for determining the optimal spatial capacity allocation of wind and PV resources in the context of national grid support. The study targets the geographic areas identified as the South African REDZs,

using the wind and solar resource datasets produced by CSIR and the EU Science Hub, respectively. The investigation involves two major aspects, namely, the development of clustered wind and PV resource maps for the target areas and the implementation and performance evaluation of various optimisation case studies to derive optimal solutions for power balance. The formulated solutions will serve as a valuable benchmark and departure point for medium- and long- term studies pertaining to the optimal deployment of wind and PV resources for increased levels of RE penetration in South Africa.

The Research Objectives (ROs) can be summarised as follows:

1) Conduct a comprehensive systematic literature review on factors influencing the formulation of a RE planning strategy to determine the optimal wind and solar capacity allocation and geographic site dispersion. Identify similar integration frameworks as well as factors which influence the current siting practices.

2) Determine the optimal wind and PV clustering process to create resource maps for the input of the integration framework.
   2.1) Determine the ideal clustering methodology: Find the optimal number of clusters as well as the clustering technique to use.
   2.2) Develop and implement a statistical resource clustering approach, which takes cognisance of the countries daily demand patterns and compare results with the commonly used temporal approach.

3) Develop the integration framework.
   3.1) Use the best-performing temporal/statistical inputted clustering algorithm to derive clustered wind and solar power geographic resource maps for the South African REDZs. The formulated clusters define the input data sets for the integration framework.
   3.2) Determine the optimal RE geospatial capacity allocation for a range of objective functions, where a risk adverse approach is adopted.

4) Achieve optimised capacity allocation inputs for the creation of realistic siting maps based on the key objectives.

5) Complete a comparative performance assessment study between the optimised allocations of wind and PV capacities versus the formulated baseline case studies, which includes the current REIPPP allocation method.

6) Determine opportunities for future work and further development of the integration framework.

1.4 Research scope

The research aims are refined to ensure concise target objectives throughout the design and implementation of each step within the study. The proposed model is depicted as a funnel implementation, as shown in Figure 1.9. The study is initially contextualised in view of South Africa's current energy climate and is then refined as the objectives are achieved in the formulation of the integration framework.

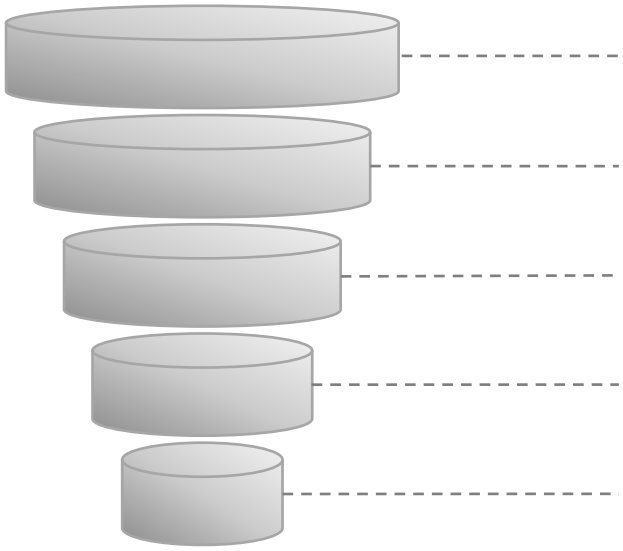


Figure 1.9: The implementation of the proposed research framework model.

1.5 Research approach and document structure

This section incorporates the research design methodology introduced by Takeda et al. [42] which is implemented in terms of the five specific research objectives coinciding with the overarching project aims. Table 1.4 describes the research objectives and the methodology, with reference to the design framework followed. Each objective is linked to the chapters in which the corresponding execution process is addressed.

Takeda *et al.* [42] introduced a method to model design processes, where design theory is implemented for the derivation of an all-inclusive model pertaining to the desired research outcomes. The Model Design Process developed by Takeda et al. can be summarised as an all-encompassing *design research cycle*, which is subdivided into five main objectives, visually shown in Figure 1.10. This model is used as the implementation approach for this research study.

The five steps are described as follows:

1) **Awareness of problem -** Gaining insight and awareness of the possible study-specific problems. Define the study-specific problem and the scope of the problem to be incorporated to determine an adequate solution.
2) **Suggestion -** Once the problem has been defined and the scope established, the possible solution theories and concepts are investigated. The appropriate methodologies and models are then selected for implementation into the development phase.
3) **Development -** The awareness and suggestion phases provide knowledge and insight into the design and development of the problem solution. Multiple development cycles can exist within this step to create an all-encompassing model.
4) **Evaluation -** The solution is evaluated, and an error feedback model is established and executed until the developed solution is ready for adoption.
5) **Conclusion -** The conclusion step initiates with the decision to adopt the evaluated solution. The results of the final model are presented, and the contributions are described.

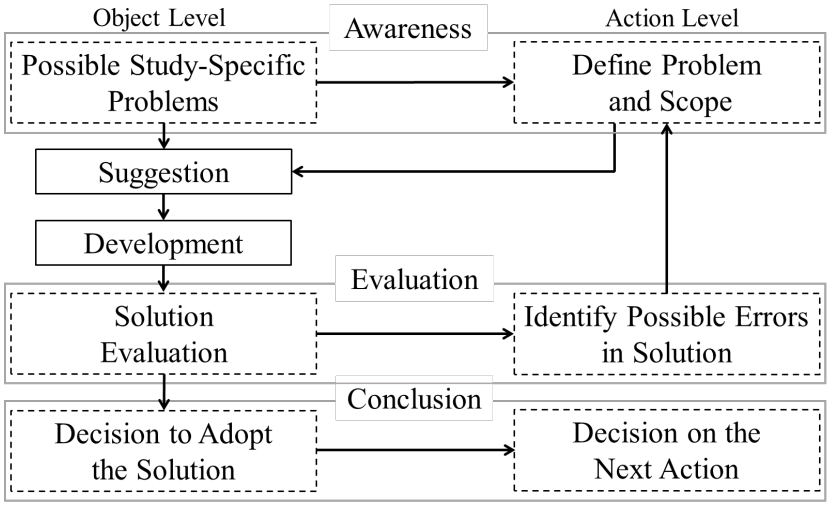


Figure 1.10: Modelling Design Processes introduced by Takeda *et al.* [42].

Table 1.4: Research structure including the design objectives and methodologies followed.

	RO	Research Objective	Chapter	Methodology
Awareness of problem	1.0	Define the generalised subject matter problem together with the research background, motivation and approach.	Chapter 1	Investigate the problem background, and as a result define the project aim, objectives, research strategy and scope.
	1.0	Understand the current energy climate within South Africa.	Chapter 1	Investigate the current renewable energy status in South Africa as well as the supporting energy framework and policy in literature.
	1.0	Determine factors influencing the formulation of a RE integration framework and examine the core competing literature.	Chapter 4	Conduct a comprehensive systematic literature review on factors influencing the formulation of a RE integration framework to determine and incentivise the optimal wind and solar capacity mix and geographic site dispersion with reference to the existing body of knowledge.
Suggestion	2.0, 2.1 & 2.2	Determine the ideal resource classification methodology. Compare statistical and temporal methods.	Chapter 3	Conduct a literature study on the performance of various clustering algorithms on temporal and statistically representative wind and solar power datasets.
	3.0 & 4.0	Determine the ideal mythologies for the development of the model.	Chapter 3 & Chapter 4	Implement and compare methodologies identified for the development of the integration framework.

		Acquire and process raw reanalysis weather and energy demand data.	Chapter 2	Raw data must be processed to obtain data sets representative of characteristic wind- and PV-system behaviour. Demand data must be processed and detrended to remove annual anomalies.
Development	3.1	Define and implement clustering objectives.	Chapter 3	Implement methodologies to create model.
	3.1	Create final clustered wind and solar power geographic resource maps.	Chapter 3	Use the resultant the wind and PV clusters to create geographic resource maps.
	3.2	Define optimisation objectives.	Chapter 4	Determine novel optimisation objectives which best suit the set objectives.
	3.2	Determine the ideal sizing and dispersion of wind and PV capacities for the key objective functions identified.	Chapter 4	Implement the ideal optimisation methodology on the wind and solar power profiles.
	4.0	Achieve allocation inputs to create realistic siting maps based on the key objectives.	Chapter 4	Implement the integration framework for various key objective case studies.
Evaluation	5.0	Evaluate the model through comparison of real-world baseline case studies.	Chapter 4	Perform a comparative performance assessment study between the optimised allocations of wind and PV capacities versus defined baseline case studies, including the current REIPPP allocation method.
Conclusion	6.0	Determine opportunities for future work and further development of the integration framework.	Chapter 5	From the literature studies conducted and the framework implemented, determine opportunities for future work and further development of the model.

1.6 Expected outcome and original contributions

The main contributions of an original nature expected from the research are as follows:

1) *Development and implementation of a novel clustering approach for classification of wind and solar resources:*
   The outcome of this work is a statistical approach to the clustering of temporospatial wind and PV power resource datasets into geospatial clusters, using statistical features defined in terms of the ToU periods associated with the demand profile.

2) *Development and implementation of a novel RE capacity allocation optimisation methodology:*
   The outcome of this work is a framework whereby wind and PV capacities are assigned to the geospatial resource clusters such that objective functions defined in terms of the residual load profile are optimised. This aspect of the work incorporates a number of novel components, including the following:
   - Identifying and evaluating the performance of an appropriate optimisation algorithm.
   - Defining a range of risk-based objective functions.
   - Proposing and implementing a strategy for accommodating real-word constraints to the capacity allocations.
   - Proposing suitable performance metrics.

1.7 Conclusion: Introduction

Chapter 1 provides a detailed overview of the research background and motivation. The structured approach is also defined, including the research aim, scope, objectives, and document structure. Finally, the novel research contributions and implications are highlighted.

Chapter 2

Data acquisition and processing

2.1 Introduction and overview

The aim of this chapter is to describe the acquisition, pre-processing and preparation of the data used to construct the integration framework. Initially, the raw wind and solar resource data sets are introduced and analysed in the context of the overarching framework. The process of converting the raw data into necessary power data is detailed. The processed data sets, including the demand profile, and the formulated ToU feature vectors are also described considering the framework objectives. The conclusion of this chapter should ascertain full confidence in the underlying datasets used in the implementation of this framework.

2.2 Research data resources and study regions

2.2.1 Research study regions

The study is implemented using two separate renewable energy resource data sets associated with the eleven REDZs proposed in phase 1 and 2 of the SEA study conducted for wind and solar energy in South Africa [43]. These zones represent distinct geographic regions within South Africa and have been demarcated for the future deployment of RE generation units, based on various environmental and socio-economic factors [44]. Figure 2.1 depicts the REDZs, which are dispersed across a wide geographic area, thereby reflecting a diverse range of climatic and topographic characteristics. Figure 2.1 incorporates a zoomed in section of the Stormberg zone. The enlarged region provides a clearer view of the individual coordinates comprising the Stormberg zone. Each of the coordinates represents a 25 km^2 area, of which the temporal wind speed and solar power data profiles are known. Table 2.1 summarises the indices associated with the various REDZs, highlighting the number of coordinates per zone as well as the applicable RE technology demarcated for each zone. The applicable RE technology is decided based on the underlying resource characteristics of each region.

The data sets represent the temporal wind speed and solar irradiance profiles for a total of 4 770 locations, distributed across the eleven REDZs. Hourly wind and solar resource data, over a five-year period from 2009 to 2013, is used to represent each spatial coordinate represented in Figure 2.1. This yields a total of N_k = 43 824 intervals for each spatial coordinate, which translates to a total of 195 893 280 observations for the wind and solar power data sets, respectively.

The REDZs were selected as the focus regions for the implementation and testing of the proposed integration study. They define a vast and diverse geographic range of potential RE integration sites, which have been pre-approved by the South African DoE based on an initial EIA. However, the proposed integration framework can easily be adapted to include South Africa in geographic entirety.

Figure 2.1: Spatial coordinates comprising each Renewable Energy Development Zone.

Table 2.1: Index summary describing the South African Renewable Energy Zones.

Name	Zone Features		
	Index	Number of coordinates	Applicable Renewable Energy Technology
Overberg	1	211	Large scale wind and PV
Komsberg	2	341	Large scale wind and PV
Cookhouse	3	288	Large scale wind and PV
Stormberg	4	467	Large scale wind and PV
Kimberley	5	372	Large scale PV
Vryburg	6	361	Large scale PV
Upington	7	497	Large scale PV
Springbok	8	593	Large scale wind and PV
Emalahleni	9	369	Large scale PV
Klerksdorp	10	307	Large scale PV
Beaufort West	11	664	Large scale wind and PV

2.2.2 Wind resource data

The wind resource data is derived using the Wind Atlas for South Africa (WASA) weather model. The WASA model applies a mesoscale downscaling method [45], together with a mesoscale modelling system. The mesoscale modelling system is also referred to as the Weather Research and Forecasting model. This mesoscale downscaling method is used to generate a temporal wind speed data set, defined for a high spatial resolution, for Southern Africa. This temporal wind speed reanalysis data set is developed by the South African CSIR and Fraunhofer Institute for Wind Energy Systems [46].

This WASA model derives wind speed data for hub heights measuring 50, 80, 100 and 150 m. The wind speed dataset spans a five-year period, from 2009 to 2013, with a spatial resolution of 25 km^2 and a temporal resolution of 15 minutes. Wind speed data at the 100 m hub height is selected and is down sampled to an hourly averaged temporal resolution for the purposes of the study.

Figure 2.2 depicts the mean wind speed map for South Africa [47], with the overlayed REDZs representing a diverse range of climatic and topographical features with excellent wind resource characteristics for the zones defined in Table 2.1.

This dataset expands the entire South African region, however for this study, the data is extracted per the eleven REDZs. Explaining the dataset mathematically, the sampling times associated with the dataset can be represented by the set $\boldsymbol{T}$, given by

$$\boldsymbol{T} = \{t_k \mid k = 1,2,3 \cdots N_k\}, \tag{2.1}$$

where t_k denotes the sampling time associated with the k^{th} time instant and N_k denotes the total number of sampling intervals. The wind speed dataset for the REDZs is represented by the set $\boldsymbol{W}$, given by the relationships

$$\boldsymbol{W} = \{\boldsymbol{W}_i \mid i = 1,2,3 \cdots N_i\}, \tag{2.2}$$

$$\boldsymbol{W}_i = \{\boldsymbol{W}_{ij} \mid j = 1,2,3 \cdots N_{ij}\} \tag{2.3}$$

and

$$\boldsymbol{W}_{ij} = \{w_{ij}(t_k) \mid k = 1,2,3 \cdots N_k\}. \tag{2.4}$$

$\boldsymbol{W}_i$ denotes the set of temporal wind speed profiles associated with the i^{th} REDZ, N_i denotes the number of REDZs, $\boldsymbol{W}_{ij}$ denotes the set of temporal wind speeds associated with the j^{th} spatial location in the i^{th} REDZ, N_{ij} denotes the total number of spatial locations associated with the i^{th} REDZ, and $w_{ij}(t_k)$ denotes the wind speed for the k^{th} sampling interval.

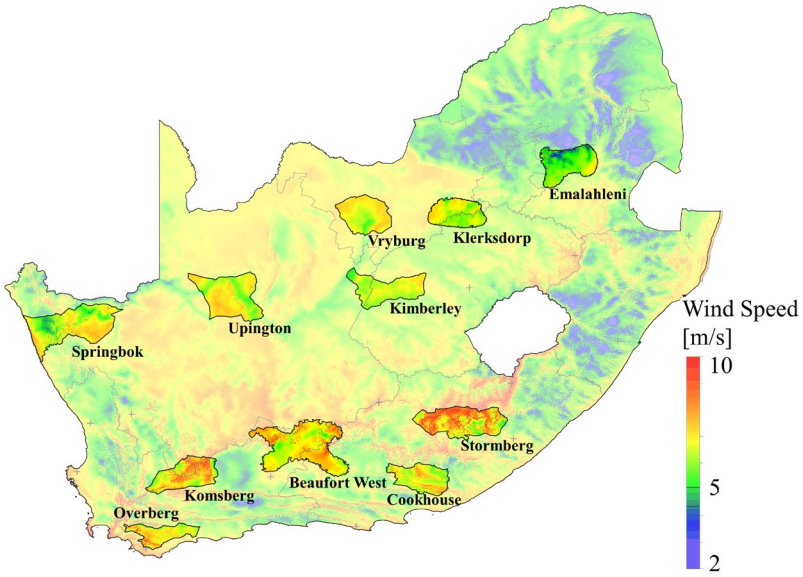


Figure 2.2: The average wind speed map for South Africa, overlayed with the Renewable Energy Development Zones.

2.2.3 Solar resource data

Radiant energy is the primary source of energy from the Sun to Earth, the measure of this energy is called solar irradiance. This is a term used to describe electromagnetic radiation which is emitted by the sun. Solar irradiance can be divided into various components, starting with Direct Normal Irradiance (DNI). This concept refers to the total solar radiation absorbed by a given surface, per unit area, which is always perpendicular to incoming rays. Diffused Radiation is the radiation which is diffused by surrounding

elements, such as atmospheric layers or particles in the sky. Lastly, reflection radiation is the radiation which is reflected by various surfaces, such as water bodies or clouds.

The dataset used for this study was developed by the Climate Monitoring Satellite Application Facility (CM SAF) and PVGIS team [48]. The dataset has an hourly temporal resolution within the annual range of 2005-2016 and a spatial resolution of 0.05 x 0.05 degrees. The dataset is labelled the PVGIS-SARAH, covering the entire African region. The dataset was derived using METEOSAT satellite imagery captured in hourly time intervals. This satellite imagery is used to estimate the cloud influence on the solar irradiation, where the cloud cover causes reflection of the incoming sunlight, which decreases the amount of radiation arriving at ground level. The satellite imagery is defined by pixels, where each individual pixel is compared at the same time each day over a month-long period and the darkest pixel seen over the measurement period is defined as clear sky. The clear sky pixel is used as a reference to effectively calculate the cloud reflectivity for all other days. The solar radiation is then calculated for clear sky conditions using the theory of radiative transfer in the atmosphere, while accounting for water vapour, ozone concentration and atmospheric aerosol density. The clear-sky model is implemented using the MAAC monthly climatology for Aerosol Optical Depth and the ERA-Interim monthly means for the water vapor. The total radiation is calculated using this clear-sky irradiance and cloud albedo [48]. Figure 2.3 depicts the mean global horizontal irradiation for South Africa, with the overlayed REDZs representing a diverse range of climatic and topographical features with high solar irradiance depicted in the northern regions.

The resultant spatiotemporal Global Horizontal Irradiation (GHI) dataset, $\boldsymbol{S}$, is defined by the relationships

$$\boldsymbol{S} = \{\boldsymbol{S}_i \mid i = 1,2,3 \cdots N_i\}, \tag{2.5}$$

$$\boldsymbol{S}_i = \{\boldsymbol{S}_{ij} \mid j = 1,2,3 \cdots N_{ij}\} \tag{2.6}$$

and

$$\boldsymbol{S}_{ij} = \{s_{ij}(t_k) \mid k = 1,2,3 \cdots N_k\}, \tag{2.7}$$

where $\boldsymbol{S}_i$ denotes the set of solar irradiance profiles associated with the i^{th} REDZ, $\boldsymbol{S}_{ij}$ denotes solar irradiance profile associated with the j^{th} spatial location in the i^{th} REDZ and $s_{ij}(t_k)$ denotes the solar irradiance for the k^{th} sampling interval.

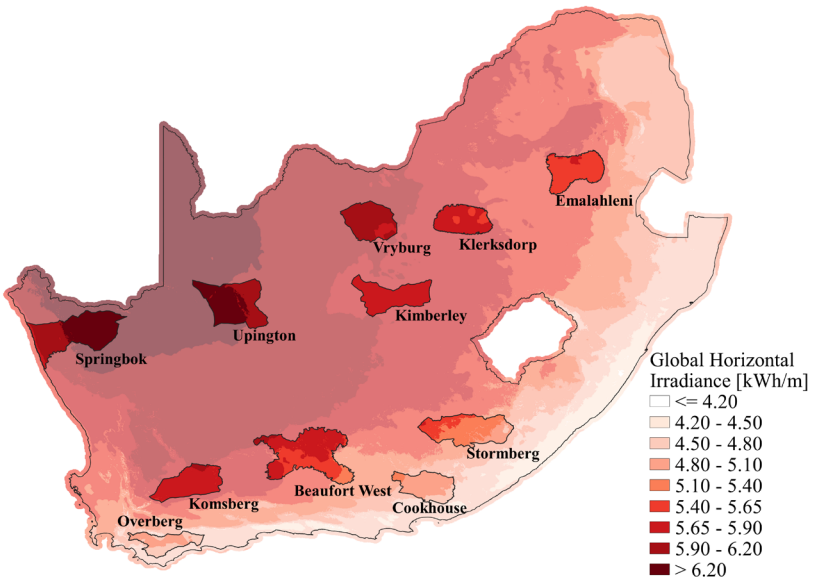


Figure 2.3: The average global horizontal irradiation map for South Africa, overlayed with the Renewable Energy Development Zones.

2.2.4 Demand profile data

The South African demand profile data available for the implementation of this study spans a 19-year period, namely from 2000 to 2019. The demand profile represents a mixture of the country's energy usage profile, bounded by the country's ability to provide the necessary demand.

In the context of the desired framework objectives, it is important to select an accurate representation of the *business as usual* case for the desired demand profile. Figure 2.4 depicts the national demand profile from 2000 to 2019. Overall, the demand profile has shown growth distortion due to generation constraints, various economic constraints and global pandemics. Since it is not necessary to use all 19 years of the demand data, it important to choose the correct annual range, which will provide the best depiction of the demand profile. Figure 2.4 shows an expected increase in demand growth between the years 2000 and 2007. Thereafter it can be seen that in 2008 the country experienced a breakdown of energy plants, which resulted in country-wide load shedding. However, from 2008 the yearly demand pattern shows a marginal decline towards 2019. This is attributed to poor economic growth, generation capacity constraints that resulted in frequent load shedding, increased embedded generation and an extensive nation-wide Demand Side Management (DSM) and Energy Efficiency (EE) program targeting the industrial, commercial and residential load sectors as well as fuel switching to Liquefied Petroleum Gas (LPG) for cooking and heating.

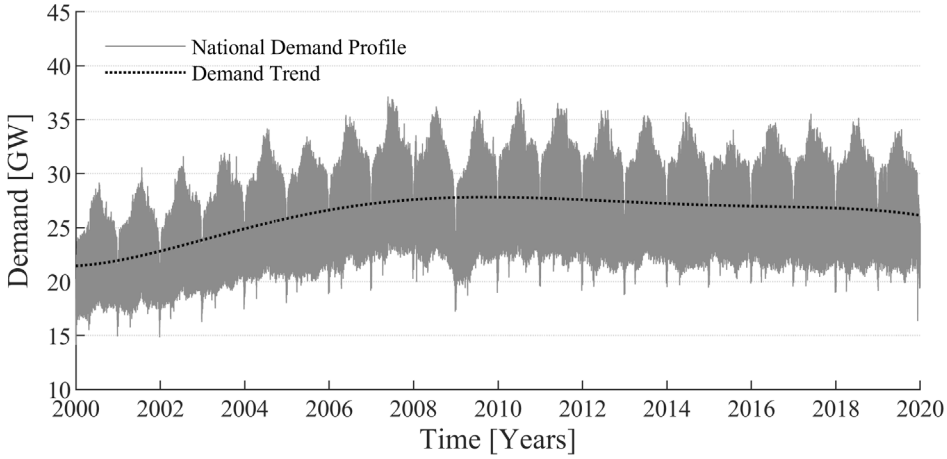


Figure 2.4: The national demand profile for South Africa, from January 2000 to December 2019.

2.3 Data conditioning and processing

Data pre-processing in context of the final framework model is multifaceted. This section describes the process followed to obtain data sets which are accurate and relevant to the framework objectives.

2.3.1 Wind resource data processing

This section will discuss two possible methods which could be used when calculating the wind power output from raw wind speed data. These methods include converting the power straight from wind speed to power via a chosen turbine class or using a multi-turbine power curve approach. All wind power calculations are based on the wind turbine power curve, which shows the general behaviour of a wind turbine under various operating wind speed conditions. This wind power curve is depicted in Figure 2.5.

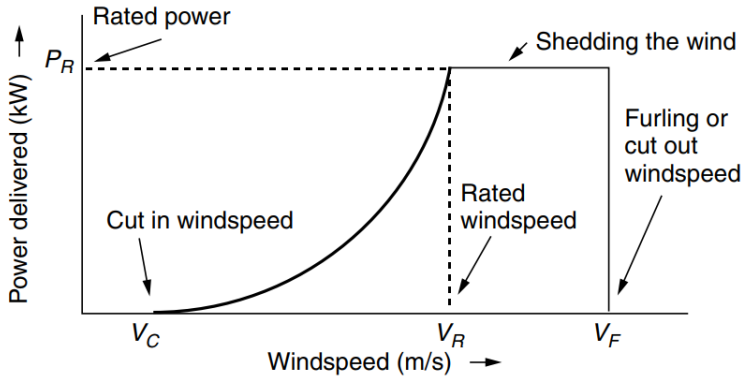


Figure 2.5: The wind turbine power curve [49].

The power output profile representative of the wind resource is locationally dependent, where the wind power profile is derived using a site-specific wind turbine power curve. Table 2.2 summarises the characteristics of the three 3.3 MW Vestas turbine models considered in the investigation. The turbine model for the j^{th} spatial location in the i^{th} REDZ is chosen based on the average wind speed for the location, $\overline{w}_{ij}$, given by

$$\overline{w}_{ij} = \frac{1}{N_k}\sum_{k=1}^{N_k} \mathrm{w}_{ij}(t_k), \qquad 2.8$$

where the temporal resolution is 15 minutes.

The wind power profiles obtained with the turbine models are normalised relative to the rating of the turbines shown in Table 2.2, i.e., 3.3 MW.

Table 2.2: Characteristics of the selected wind turbine models [50].

Average Wind Speed at Site [m/s]	Wind Turbine Model	Model Class
6.0 - 7.5	V126 - 3.3 MW	IEC IIIA
7.5 - 8.5	V117 - 3.3 MW	IEC IIA
8.5 - 10.0	V112 - 3.3 MW	IEC IB

To investigate the effects of aggregation of wind speed over a finite area, the multi-turbine power curve approach, proposed by Nørgaard and Holttinen [51], is investigated. In this methodology the temporal, location specific wind speed profile $\boldsymbol{W}_{ij}$ is block-averaged using a sliding window to filter the wind speed fluctuations associated with the propagation of the wind over the target wind farm area. The block-averaged wind speed for the k^{th} sampling interval, $\mathrm{w}_{ij}^{F}(t_k)$, is expressed by the mathematic relationship [51]

$$\mathrm{w}_{ij}^{F}(t_k) = \frac{1}{N_{ij}+1}\sum_{m=k-\frac{N_{ij}}{2}}^{k+\frac{N_{ij}}{2}} \mathrm{w}_{ij}(t_m), \tag{2.9}$$

where $w_{ij}(t_m)$ denotes the original wind speed in $\boldsymbol{W}_{ij}$ at time t_m and $N_{ij}+1$ denotes the size of the averaging window as several sampling intervals. N_{ij} is an even-number integer derived from the relationship [51]

$$N_{ij} = \frac{D}{\bar{w}_{ij}\Delta t}, \tag{2.10}$$

where D denotes the spatial dimension of the target area, $\overline{\mathrm{w}}_{ij}$ denotes the mean wind speed of the original time series and Δt denotes the temporal resolution of the original time series. The spatial resolution of the wind resource dataset, D, is equal to 5 km.

The multi-turbine power curve, m_{ij}, for a site is derived by convolving the single turbine power curve selected for the site, p_{sij}, and the spatial wind speed distribution of the site, n_{sij}. This is defined by the relationship [51]

$$m_{ij} = p_{sij} * n_{sij}. \tag{2.11}$$

A normal distribution is assumed for the spatial wind speed distribution describing the wind speeds for the individual wind turbines at the wind farm. The standard deviation of this distribution is determined from parameterised curves using the spatial dimension, D, and turbulence intensity of the site [51].

The block averaged wind speed is applied to the multi-turbine power curve to produce a per unit power time series using the relationship [51]

$$\mathrm{P}_{ij}(t_k) = m_{ij}(\mathrm{w}_{ij}^{F}(t_k)). \tag{2.12}$$

The mean of the spatial wind speed distribution is initially set to zero, and is adjusted iteratively until

$$\sum \mathrm{P}_{ij}(t_k) = \sum \mathrm{P}_{sij}(t_k), \tag{2.13}$$

where

$$\mathrm{P}_{sij}(t_k) = p_{sij}\left(\mathrm{w}_{ij}^{F}(t_k)\right). \quad 2.14$$

The multi-turbine power curve introduces a low pass filtering effect on the wind power profiles. The results show that this effect is relatively marginal for the geospatial resolution of 5 km by 5 km that applies for the input dataset, and the results are therefore presented for the case study where this aggregation is neglected.

The resultant set of wind power profiles $\boldsymbol{P}^{w}$ is defined by the relationships

$$\boldsymbol{P}^{w} = \{\boldsymbol{P}_{i}^{w} \mid i = 1,2,3 \cdots N_i\}, \quad 2.15$$

$$\boldsymbol{P}_{i}^{w} = \{\boldsymbol{P}_{ij}^{w} \mid j = 1,2,3 \cdots N_{ij}\} \quad 2.16$$

and

$$\boldsymbol{P}_{ij}^{w} = \{p_{ij}^{w}(t_k) \mid k = 1,2,3 \cdots N_k\}, \quad 2.17$$

where $\boldsymbol{P}_{i}^{w}$ denotes the set of wind power profiles associated with the i^{th} REDZ, $\boldsymbol{P}_{ij}^{w}$ denotes the wind power profile associated with the j^{th} spatial location in the i^{th} REDZ and $p_{ij}^{w}(t_k)$ denotes the wind power for the k^{th} sampling interval.

2.3.2 Solar resource data processing

The relative power output of a PV plant is proportional to the solar irradiance found at the Earth's surface, for a specific area. A more accurate power output occurs when considering the tilted irradiance, together with the ambient temperature at each solar plant. In this case, the direct tilted irradiance is that which hits the solar panel on the suns direct path, the diffuse tilted irradiance is that which has been scattered by atmospheric molecules and particles, and the reflected tilted irradiance is that which is reflected off other objects in the panels surrounding area.

The spatiotemporal solar irradiation profiles are translated to power profiles using a fixed 1 kWp crystalline silicon panel, with an azimuth angle such that the panel faces north, and the tilt angle is determined based on location for maximum annual yield. Power conversion is performed using the methodology defined by Huld et al. [52], which is a variant of King's model [53], with the system losses set to 14 %. This approach takes cognisance of the mounting type, ambient temperature and wind patterns relative to the PV module.

The power measured from a PV module under Standard Test Conditions (STC) is known as the nominal power of the module. This STC is dictated by the international standard IEC-60904-1. The standard testing conditions state that the light intensity over the entire module surface should be G_{STC} = 1000 $\left[\frac{W}{m^2}\right]$ and the operating temperature should be T_{STC} = 25 °C. The spectrum of light should comply with the global spectrum set in IEC 60904-3. The normal operating conditions are however largely influenced by external factors, which should be accounted for when estimating the power output of an operational PV plant. The power is dependent on the effects of the module temperature T_M and the irradiance G, defined mathematically as [52]

$$P(G',T') = G'(P_{STC,m} + k_1 \ln(G') + k_2 \ln(G')^2 + k_3 T' + k_4 T' \ln(G') + k_5 T' \ln(G') + k_6 T'^2), \quad 2.18$$

where

$$G' = \frac{G}{G_{STC}} \tag{2.19}$$

and

$$T'_M = T_M - T_{STC}. \tag{2.20}$$

The coefficients used in Equation (2.18) are defined based on measurements preformed on the PV module technology used, in this case a crystalline silicon (c-Si) panel, where [52]

$$k_1 = -0.017237,\ k_2 = -0.040465,\ k_3 = -0.004702,\ k_4 = 0.000149,\ k_5 = 0.000170\ and\ k_6 = -0.000005. \tag{2.21}$$

The module temperature T_M is affected by both the ambient temperature and the wind effects surrounding the PV module. The resultant temperature value is therefore calculated as [54]

$$T_M = T_a - \frac{G}{(U_0 + U_1 W)}, \tag{2.22}$$

where T_a is the ambient temperature, W denotes the wind speed, U_0 denotes the coefficient describing the effect of the radiation on the module temperature and U_1 denotes the coefficient describing the cooling effect caused by the wind [54, 55].

This mathematical process is used to calculate the power output per module, thereafter multiple power losses must be accounted for, namely, DC-AC transformation losses, cable losses as well as panel degradation losses accumulated over time. These losses are complex and are usually summated to a total user specific system loss percentage value, where a 14 % loss percentage value is recommended.

The resultant solar power profile set, $\boldsymbol{P}^s$, is defined by the relationships

$$\boldsymbol{P}^s = \{\boldsymbol{P}_i^s \mid i = 1,2,3 \cdots N_i\}, \tag{2.23}$$

$$\boldsymbol{P}_i^s = \{\boldsymbol{P}_{ij}^s \mid j = 1,2,3 \cdots N_{ij}\} \tag{2.24}$$

and

$$\boldsymbol{P}_{ij}^s = \{p_{ij}^s(t_k) \mid k = 1,2,3 \cdots N_k\}, \tag{2.25}$$

where $\boldsymbol{P}_i^s$ denotes the set of solar power profiles associated with the i^{th} REDZ, $\boldsymbol{P}_{ij}^s$ denotes the solar power profile associated with the j^{th} spatial location in the i^{th} REDZ and $p_{ij}^s(t_k)$ denotes the solar power for the k^{th} sampling interval. The resulting solar power dataset has an hourly resolution.

2.3.3 Demand profile data processing

The demand profile used in the study is derived from the South African national demand profile for the period from January 2010 to December 2014. This excludes the period of strategic growth from 2000 to 2007. However, it is recognised that strategic growth may have to be included in future medium- to long-term studies. This period also excludes the annual profile for 2015, which differs from the surrounding profiles, shown in Figure 2.4.

Figure 2.6 depicts the selected demand profile range, which is detrended using a linear least squares regression approach with the view to remove year on year trends, thereby focusing on the daily and seasonal characteristics. The detrended profile is subsequently normalised to the peak value P_{max}^d, where

P^d_{max} = 36.3713 GW for this period. This yields the profile depicted in Figure 2.7, which is used in the framework development. The normalised demand dataset $\boldsymbol{P}^d$ is defined by the relationship

$$\boldsymbol{P}^d = \{p^d(t_k) \mid k = 1,2,3 \cdots N_k\}, \tag{2.26}$$

where $p^d(t_k)$ denotes the normalised demand value for the k^{th} sampling interval.

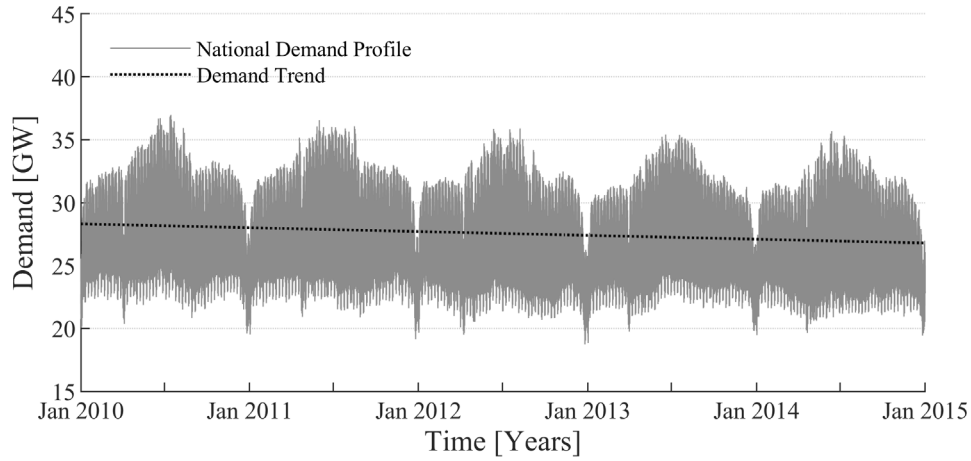


Figure 2.6: The demand profile for South Africa with the annual linear trend overlay.

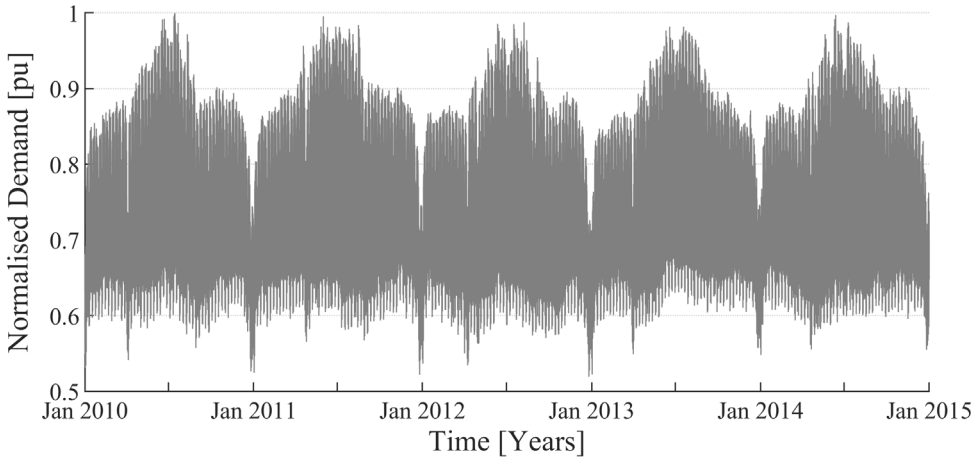


Figure 2.7: The detrended, normalised demand profile for South Africa.

2.3.4 Resource feature vector formulation

In the highly regulated South African electricity market, the national power utility implements fixed bulk-supply ToU tariff structures for large consumers and resellers connected at the transmission level [37]. The associated ToU periods and electricity costs are based on the national demand profile and cost of generation and transmission for the power network.

For the implementation of the proposed framework, a novel method for quantifying temporal resource profiles is introduced. The approach defines the spatiotemporal resource profiles as a set of feature vectors, based on the statistical properties defined within the daily and seasonal ToU periods.

This bulk supply tariff structure features two distinct demand seasons, namely, a low demand season ranging from the 1st of September to the 31st of May, and a high demand season ranging from the 1st of June to 31st of August. Furthermore, the tariff structure incorporates three daily ToU categories, namely,

off-peak, standard, and peak periods. The three daily ToU periods can be further subdivided into the daily temporal composition defined by the Megaflex tariff structure, displayed in Figure 2.8. Figure 2.8 summarises the seasonal demand as well as the weekday ToU periods for the South African Megaflex ToU tariff structure, while Table 1.3 summarises the normalised cost of energy for the demand periods [37]. There is significant differentiation in the cost of energy between the daily ToU periods, and between the low demand and high demand seasons, respectively. It is therefore proposed that a variable resource such as wind energy be classified and quantified in the context of these ToU periods.

Figure 2.9 and Figure 2.10 depict histograms of the hourly weekday demand values for the low demand and high demand seasons, respectively. The histograms show that for the high demand season, the loads in the peak periods are higher and occur more frequently compared to the low demand season. The demand periods depicted by the histograms in Figure 2.9 and Figure 2.10 are normalised to the number of sampling intervals within each demand season over the five-year span.

The ToU feature vectors derived for the spatiotemporal power resource profiles consists of the mean, μ, and standard deviation, σ, of the normalised resource for each of the daily ToU period categories depicted in Figure 2.8. The feature vector can be categorised using the three ToU periods classified in Table 2.3 as well as further expanding the vector to include the daily temporal split as shown in Figure 2.8 and Table 2.4.

Defining the feature vector which comprises of the three ToU periods (peak, standard and off-peak) for each demand season (high and low demand), yields a total of 6 ToU periods, comprising of 2 features per period. This results in a total of 12 defining features per spatial location, 6 features for the low demand season and 6 features for the high demand season. The feature vector can be represented mathematically by the set

$$\boldsymbol{F^6} = \{\boldsymbol{F_L^6}, \boldsymbol{F_H^6}\}, \tag{2.27}$$

where

$$\boldsymbol{F_L^6} = \{\mu_{OP}^L, \sigma_{OP}^L\}, \{\mu_S^L, \sigma_S^L\}, \{\mu_P^L, \sigma_P^L\} \tag{2.28}$$

and

$$\boldsymbol{F_H^6} = \{\mu_{OP}^H, \sigma_{OP}^H\}, \{\mu_S^H, \sigma_S^H\}, \{\mu_P^H, \sigma_P^H\}. \tag{2.29}$$

Defining the feature vector which comprises of all daily ToU periods yields a total of 13 ToU periods, 7 in the low demand season and 6 in the high demand season, as shown in Figure 2.8. Each of the 13 ToU periods is defined by the two statistical features, yielding a total of 26 defining features per spatial location, i.e., 14 features for the low demand season and 12 features for the high demand season. The feature vector can be represented mathematically by the set

$$\boldsymbol{F^{13}} = \{\boldsymbol{F_L^{13}}, \boldsymbol{F_H^{13}}\}, \tag{2.30}$$

where

$$\begin{aligned}\boldsymbol{F_L^{13}} = {} & \{\mu_{OP_M}^L, \sigma_{OP_M}^L\}, \{\mu_{S_M}^L, \sigma_{S_M}^L\}, \{\mu_{P_M}^L, \sigma_{P_M}^L\}, \{\mu_{S_{Mi}}^L, \sigma_{S_{Mi}}^L\}, \\ & \{\mu_{P_E}^L, \sigma_{P_E}^L\}, \{\mu_{S_E}^L, \sigma_{S_E}^L\}, \{\mu_{OP_E}^L, \sigma_{OP_E}^L\}\end{aligned} \tag{2.31}$$

and

$$\begin{aligned}\boldsymbol{F_H^{13}} = {} & \{\mu_{OP_M}^H, \sigma_{OP_M}^H\}, \{\mu_{P_M}^H, \sigma_{P_M}^H\}, \{\mu_{S_{Mi}}^H, \sigma_{S_{Mi}}^H\}, \\ & \{\mu_{P_E}^H, \sigma_{P_E}^H\}, \{\mu_{S_E}^H, \sigma_{S_E}^H\}, \{\mu_{OP_E}^H, \sigma_{OP_E}^H\}.\end{aligned} \tag{2.32}$$

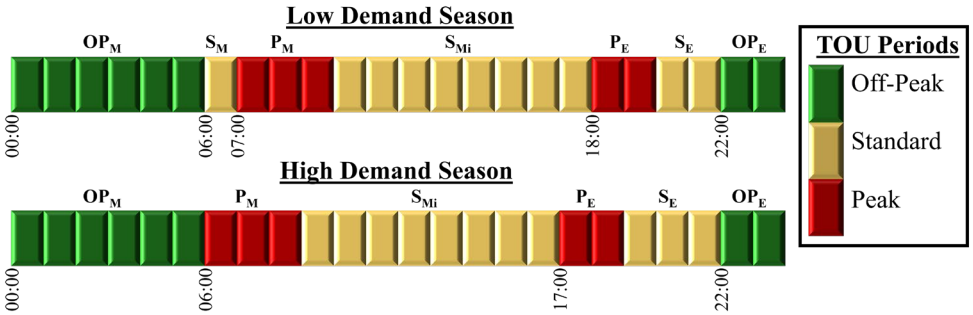


Figure 2.8: Megaflex ToU tariff structure showing the weekday peak (red), standard (yellow) and off-peak (green) periods for the high and low energy demand seasons, respectively [37].

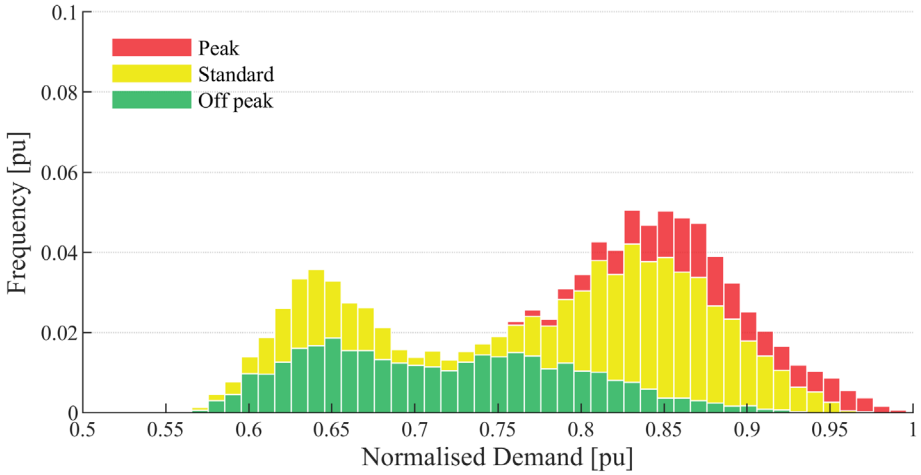


Figure 2.9: High demand season histogram, depicting the ToU period breakdown.

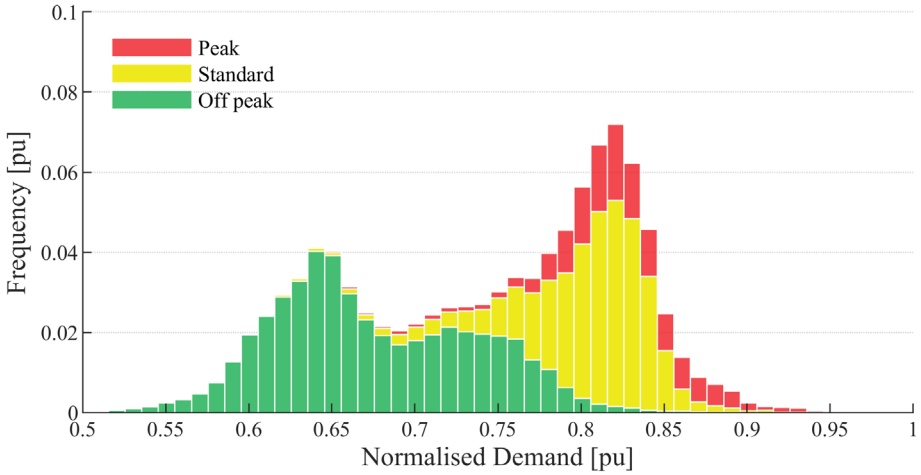


Figure 2.10: Low demand season histogram, depicting the ToU period breakdown.

Table 2.3: Statistical ToU features vector parameters for the 6 ToU periods.

ToU Period	Period Times	Features
	Low Demand Season	
Off-peak	00:00 - 06:00, 22:00 - 24:00	$\mu_{OP}^{L}, \sigma_{OP}^{L}$
Standard	06:00 - 07:00, 10:00 - 18:00, 20:00 - 22:00	$\mu_{S}^{L}, \sigma_{S}^{L}$
Peak	07:00-10:00, 18:00-20:00	$\mu_{P}^{L}, \sigma_{P}^{L}$
	High Demand Season	
Off-peak	00:00 - 06:00, 22:00 - 24:00	$\mu_{OP}^{H}, \sigma_{OP}^{H}$
Standard	09:00 - 17:00, 19:00 - 22:00	$\mu_{S}^{H}, \sigma_{S}^{H}$
Peak	06:00 - 09:00, 17:00 - 19:00	$\mu_{P}^{H}, \sigma_{P}^{H}$

Table 2.4: Statistical ToU features vector parameters for the 13 ToU periods for the Megaflex tariff.

Low Demand Season		High Demand Season	
ToU Period	Features	ToU Period	Features
Morning off-peak 00:00 - 06:00	$\mu_{OP_M}^{L}, \sigma_{OP_M}^{L}$	Morning off-peak 00:00 - 06:00	$\mu_{OP_M}^{H}, \sigma_{OP_M}^{H}$
Morning standard 06:00 - 07:00	$\mu_{S_M}^{L}, \sigma_{S_M}^{L}$	Morning peak 06:00 - 09:00	$\mu_{P_M}^{H}, \sigma_{P_M}^{H}$
Morning peak 07:00 - 10:00	$\mu_{P_M}^{L}, \sigma_{P_M}^{L}$	Midday standard 09:00 - 17:00	$\mu_{S_{Mi}}^{H}, \sigma_{S_{Mi}}^{H}$
Midday standard 10:00 - 18:00	$\mu_{S_{Mi}}^{L}, \sigma_{S_{Mi}}^{L}$	Evening peak 17:00 - 19:00	$\mu_{P_E}^{H}, \sigma_{P_E}^{H}$
Evening peak 18:00 - 20:00	$\mu_{P_E}^{L}, \sigma_{P_E}^{L}$	Evening standard 19:00 - 22:00	$\mu_{S_E}^{H}, \sigma_{S_E}^{H}$
Evening standard 20:00 - 22:00	$\mu_{S_E}^{L}, \sigma_{S_E}^{L}$	Evening off-peak 22:00 - 24:00	$\mu_{OP_E}^{H}, \sigma_{OP_E}^{H}$
Evening off-peak 22:00 - 24:00	$\mu_{OP_E}^{L}, \sigma_{OP_E}^{L}$		

2.4 Conclusion: Data acquisition and processing

Chapter 2 provides a detailed overview of the study regions and data sets used. The data processing methods are thoroughly described to ascertain full confidence in the resulting data sets used in the implementation of this framework.

Chapter 3

Clustering of renewable energy resource data

3.1 Overview

This section details the RE resource clustering method followed, which is implemented as the pre-processing step to derive the input of the optimisation framework. The aim is to produce clustered wind and solar resource maps, which best represent the geographic areas for the underlying features and grid support capabilities. This section also entails a literature review of various clustering approaches used in similar RE resource clustering scenarios. A comparative study is implemented on the performance of various clustering algorithms using both temporal profiles and statistical distribution profiles derived from the wind and PV data sets.

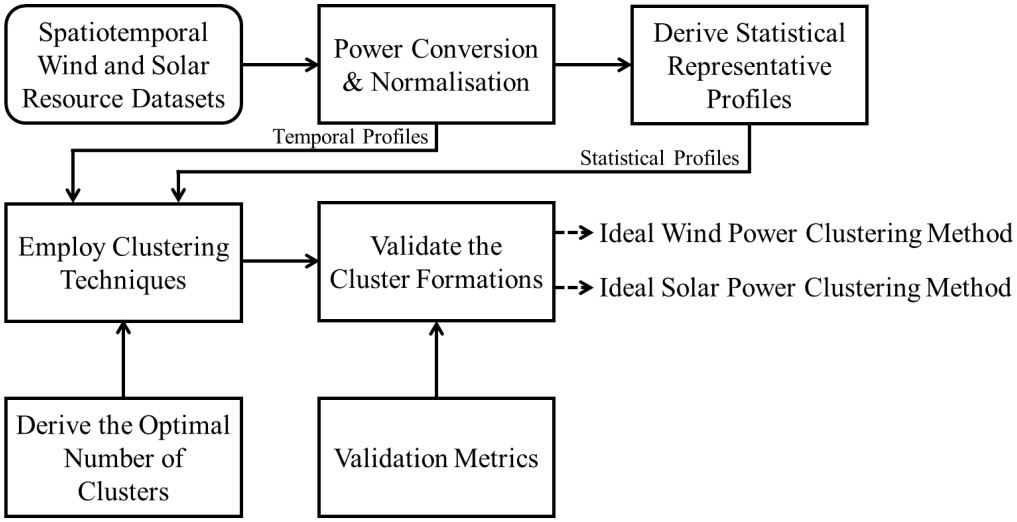


Figure 3.1: Flow diagram depicting the process for determining the ideal wind and solar power clustering methodologies.

3.2 Literature review

The rapid growth in global RE penetration levels has given rise to increased interest in strategies for mitigating the associated short-term to medium-term operational grid integration risks. The characterisation and optimal allocation of RE resources are of paramount importance in this context. Resource clustering and classification have received considerable attention in literature [56, 57, 58, 59]. However, limited work has been published on the use of resource clustering as an initial step for the optimal geographic allocation of RE generation capacity in medium-term to long-term grid integration studies.

South Africa represents an example of an emerging market in the initial transition stage from a predominately fossil fuel-based energy scenario to an integrated generation fleet with a high penetration of RE. however, at present the current energy mix shows limited wind and PV resource integration [14]. Therefore, appropriate resource assessment and grid integration studies are of major importance to ensure a successful transition to the large-scale integration of RE.

In the absence of appropriate planning and resource management, high penetration levels of RE may lead to substantial undersupply or oversupply in certain demand periods [3, 4]. The demand profile exhibits well-defined daily and seasonal ToU periods, which translate into ToU tariff systems for large consumers and resellers. The planning and optimisation of RE resources, in the context of grid support, is therefore particularly important in ensuring a successful migration to a sustainable RE integration scenario.

The optimised geographic allocation of large-scale RE resources requires meteorological data sets with sufficient geographic coverage and temporal resolution. Spatiotemporal wind and solar resource datasets with high temporal and spatial resolutions are readily available for the South African geographic region. However, the practical use of these large datasets for resource analysis and grid integration studies requires substantial analytical effort and computational power. These challenges can be mitigated by implementing data reduction techniques, such as clustering the constituent profiles into archetypical spatiotemporal categories [60]. Furthermore, clustered spatiotemporal resource data sets facilitate high-level interpretation of the resource characteristics in the context of applications such as optimised geospatial capacity allocation [59], resource adequacy, forecasting and optimal dispatch [60].

Clustering of spatiotemporal data has been applied extensively in the context of RE resource classification, especially for wind resources. Consequently, the review of related works examines the application of clustering techniques implemented for the classification of RE resource data sets.

Burlando et al. [58] implemented various clustering approaches, using combinations of agglomerative clustering methods and distance metrics, to classify mesoscale wind speed profiles. The data utilised in the investigation was collected from 11 weather stations located along the perimeter of the island of Corsica, for a period of three calendar years. This study yielded three distinct anemological regions, and 8 regional wind regimes for the geographic area. It is shown that the clustered anemological regions correlate with established climatic zones of the island. Moreover, the derived wind regimes correlate with synoptically driven and thermally forced wind patterns in the area.

Cassola et al. [59] proposed a methodology to minimise the cumulative variability of wind energy by optimising the location of wind energy plants. The study was implemented using wind data collected from 10 weather stations for a period of 3 calendar years for the island of Corsica. The data is clustered and mean-variance optimisation is performed on the clustered and unclustered datasets. The results from both optimisations show similar performance, which suggests that clustering is a feasible data reduction procedure for wind energy siting applications.

Leenman and Phillipson [61] employed clustering as an initial data reduction step to derive several partitioned wind zones, before implementing a methodology for the optimal allocation of wind power capacity in the Netherlands to reduce transmission line losses. The study used wind data from 50 meteorological stations, over a period of 10 years. The results of this investigation show a significant relation between wind farm siting and transmission line losses.

Snel and Lundstedt [62] applied clustering on 25 years of solar wind data, using a Self-Organising Map (SOM). The study yielded clusters that are representative of events such as coronal mass ejections and solar holes. Berkovic [63] implemented a SOM clustering approach to derive distinct wind regimes for Israel using approximately 5 years of 10-minute averaged wind measurements from 53 meteorological stations. The resulting clusters showed high correlation with subjective and quantitative analyses in prior works.

Yesilbudak [64] applied k-means clustering to partition monthly averaged wind speed data for 75 regions in Turkey. The performance of the squared Euclidean distance, Manhattan distance, Cosine similarity and Pearson distance metrics are compared in the investigation. It is concluded that the clustering results are useful for identifying the provinces with the best potential for wind energy generation. Furthermore,

based on silhouette coefficients, it is shown that the squared Euclidean metric delivers the best clustering performance.

Pinto et al. [65] proposed a methodology to classify archetypal weather developments to assess potential variations in storm impacts in future decades using reanalysis data for the state of North Rhine-Westphalia, Germany. The core of the methodology is centred on the use of principal component analysis and the k-means algorithm to classify the synoptic development conditions. The study yielded results that are consistent with those presented in similar studies.

Dong [66] applied k-means clustering to group monthly weather patterns based on similarity, using daily reanalysis data spanning a 38-year period for the Southern Plains of the United States. The formulated clusters represent the typical weather regimes, where the results are assessed using a baseline frequency of occurrence of large-scale atmospheric circulations. The results show well formulated wind resource clusters using the novel approach when compared with the generic approach of subjectively assessing daily weather maps.

Nahmmacher et al. [67] applied the hierarchical clustering algorithm for the grouping of historical days of the year displaying similar weather patterns. The number of clusters chosen is determined by calculating the sum of squared error between cluster centroids and all cluster members as a function of the number of clusters, otherwise known as the elbow point method. The daily clusters are obtained as the initial step in deriving and selecting representative days as power system model inputs.

Overall, these studies support the concept of implementing clustering as an initial dimensionality reduction step to decrease the computational burden associated with optimisation studies involving the large-scale grid integration of RE.

Temporal wind resource profiles are generally clustered based on averaged, unweighted similarity metrics [58, 61, 64], but it is recognised that L_p norm distance metrics, such the Euclidean distance and Root Mean Square Error (RMSE), can produce poor performance for high dimensionality time-series clustering [60]. The importance of relatively short daily demand periods of high interest to system operations, such as peak periods, is furthermore poorly represented when clustering spatiotemporal wind power profiles using averaged-based similarity metrics. This can result in clusters with good similarity in terms of the overall time series observations but are sub-optimally constituted in the context of operational grid support criteria.

Resource dataset clustering techniques are performed within this study to identify a set of characteristically similar geographic areas [68]. The clustering methodologies can be implemented for various, user specified, defining characteristics. Each clustering technique, regardless of the input factors, requires a predefined similarity/dissimilarity measure. Clustering also requires the derivation of the optimal number of clusters which most accurately describe the subdivided dataset.

3.3 Methodology review

This section compares various clustering methodologies to identify the method that best segregates and classifies the RE resource dataset, while reducing the data processing complexity and storage space required. This section will specifically focus on the clustering of raw and converted power time-series data as well as the statistical and model-based features extracted from the time-series data sets.

The clustering methodologies compared include k-means clustering, partition around medoids (PAM) [69, 70], clustering large applications algorithm (CLARA) [71], fuzzy C means clustering [72, 73], agglomerative and divisive analysis clustering, as well as clustering using SOMs [74, 75].

A time series represents a set of dynamic observations which change chronologically as a function of time. This data is generally referred to as temporal data, which is often high dimensional data [76, 77]. When working with a temporal dataset, it is important to identify the similarity measures which are most appropriate for the chosen application. Since temporal data is usually noisy and includes multiple outliers, selecting the correct similarity measures is often a major challenge [78]. The whole sequence matching technique is a commonly used approach, which allows for the consideration of whole time series lengths during distance calculations [79]. Another approach is subsequence clustering, where a shorter segment of the whole time series is selected for clustering implementation.

There are several clustering methods which identify different patterns within the data. These approaches can include recognising the dynamic changes or correlation patterns in the temporal data [80], as well as prediction and recommendation patterns based on the raw time series, often used in forecasting applications [81, 82, 83].

In whole sequence clustering, three different approaches can be adopted, namely, the shape, feature or model-based approach. The shape-based approach is mostly used on the raw time series data, as its aim is to best match the shape between various observations within the time series. The feature-based approach is implemented on feature vectors which have been extracted as a representative of the original raw time series. This can be quantified as the statistical representation of the raw time series, rather than a temporal dataset. In the model-based approach, the raw time series is represented as a set of model parameters for each observation. However, this method shows a reduction in performance when the time series observations are close in proximity [84].

Before the clustering methodologies can be applied on the time series dataset, the concept of dimension reduction may be applied [85, 86, 87]. Dimension reduction reduces the memory allocation requirements. Computational complexity is also reduced when implementing the distance metric, which in turn increases the speed of the algorithmic clustering process [88]. Choosing the correct reduction methodology often determines the balance between computational complexity and representative data quality [89]. A large amount of research has been done on various time series dimensionality reduction techniques [90, 91, 88, 92, 93]. A common factor in these techniques is correctly selecting a method which ensures optimal end results.

There are various similarity measures used on time series data, with the most common metrics being dynamic time warping, Pearson correlation and the Euclidean distance [94, 95, 96, 97].

The Euclidean distance, $d_{euc}(x, y)$, is defined by the relationship [98]

$$d_{euc}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\,, \tag{3.1}$$

where x and y denote two vectors of length n within the dataset.

The Pearson correlation distance, $d_{cor}(x, y)$, is defined by the relationship [98]

$$d_{cor}(x, y) = 1 - \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{3.2}$$

where $\bar{x}$ and $\bar{y}$ denote the means of x and y, respectively. This correlation distance measures the linear relationship degree between elements x and y.

These distance metrics are selected based on approach, where the main approaches include finding similar series dataset instances based on time, shape or change in the dataset. The time-based approach focuses on the correlation of each time series for each time step, where examples of this approach include Fourier transforms or Piecewise Aggregate Approximation [99]. Similarity based on shape is a similarity

metric which is independent of the time points and focuses mainly on the similarities in data patterns. Research shows that this method displays enhanced results when compared with the time-based method [100].

3.3.1 Clustering algorithms

The clustering algorithms applied and compared in this study include the k-means algorithm, PAM, CLARA, the Agglomerative Hierarchical algorithm, a model-based clustering algorithm and the SOM approach.

When formulating the clustering algorithms mathematically, the un-clustered dataset, P, can be represented by the expression [101]

$$\boldsymbol{P} = \{p_i,\ i = 1, 2, 3, \dots, N^i\}, \tag{3.3}$$

where p_i denotes the i^{th} element and N^i denotes the number of observations in the set. The set of clusters, C, can be represented by the expression

$$\boldsymbol{C} = \{\boldsymbol{C_j}|\boldsymbol{C_j} \subset \boldsymbol{P},\ j = 1, 2, 3, \dots, N^j\}, \tag{3.4}$$

where $\boldsymbol{C_j}$ comprises of elements within the set $\boldsymbol{P}$, and it denotes the set of N^j clusters.

The set of observations within a cluster, $\boldsymbol{C_j}$, can be represented by the expression

$$\boldsymbol{C_j} = \{C_{jk}|C_{jk} \subset\ P,\ j = 1, 2, 3, \dots, N^{jk}\}, \tag{3.5}$$

where C_{jk} and N^{jk} denote the k^{th} observation within the j^{th} cluster and the number of observations, respectively in cluster set C_j. The set of centroids associated with the clusters, $\boldsymbol{W}$, is represented by the expression

$$\boldsymbol{W} = \{W_j|W_j \in\ C_j,\ j = 1, 2, 3, \dots, N^j\}, \tag{3.6}$$

where W_j denotes the j^{th} centroid [101].

In the k-means partitioning clustering algorithm each observation of the set $\boldsymbol{P}$ is iteratively assigned to a cluster $\boldsymbol{C_j}$ with a characteristically similar centroid, W_j. The initially appointed cluster assignments of each element, p_i, remains unchanged until convergence occurs. The number of clusters for the k-means method must be predefined and methods such as the elbow point method and information criterion approach can be used to determine this predefined number [102, 103].

The CLARA algorithm is based on the PAM method, which has been adapted for use on large datasets [104]. The quality of the selected medoids is determined by the average dissimilarities between each clustered point and its corresponding medoid. This quality measure is identified as the medoid rating function. The set of medoids, $\boldsymbol{M}$, can be represented by the mathematical set expression

$$\boldsymbol{M} = \{m_j,\ j = 1, 2, 3, \dots, N^j\}, \tag{3.7}$$

where m_j denotes the j^{th} medoid [101]. The rating function, $R(\mathrm{m_j}, C_{jk})$, is defined by the mathematical relationship [104]

$$R(m_j, C_{jk}) = \sum_{C_{jk}} \frac{d\left(C_{jk}, rpst(m_j, p_i)\right)}{N^{jk}}, \tag{3.8}$$

where $d\left(C_{\mathrm{jk}}, \mathrm{rpst}(m_j, p_i)\right)$ represents the dissimilarity between two dataset elements C_{jk} and $rpst(m_j, p_i)$ and $rpst(m_j, p_i)$ represents the medoid closest to an element p_i. By using this measure after suitable iterations, the cluster with the smallest dissimilarity sum is retained.

The agglomerative clustering method is often referred to as a *bottom-up* approach, whereby each element is initialised as its own cluster. Thereafter, similar clusters are merged with successive levelling within the process, starting from the bottom of the dendrogram until the top. The dendrogram is a tree-like structure that represents all elements within the dataset. This iterative process continues until one large cluster, representing the entire dataset, is reached [105].

Model-based clustering, classification, and density estimation is based on Gaussian mixture modelling. In the un-clustered set of $\boldsymbol{P}$, the distribution of each element is represented by a probability density function through a finite mixture model of p_i components. The probability density function, $f(p_i, \psi)$, is represented by the mathematical expression [106]

$$f(p_i, \psi) = \sum_{k=1}^{P} \pi_k f_k(p_i, \theta_k), \tag{3.9}$$

where ψ represents the mixture model parameters, given by

$$\psi = \{\pi_1, \dots, \pi_{P-1}, \theta_1, \dots, \theta_P\}. \tag{3.10}$$

$f_i(p_i, \theta_k)$ denotes the k^{th} component density for the observation p_i, where the parameter vector and θ_k,($\pi_1, \dots, \pi_{P-1}$) are the mixing probabilities, where

$$\pi_k > 0 \tag{3.11}$$

and

$$\sum_{k=1}^{P} \pi_k = 1. \tag{3.12}$$

The Gaussian finite mixture model is fitted by an Expectation Maximisation (EM) algorithm. The EM algorithm is used in conjunction with statistical models, whereby it iteratively finds the maximum possibility of parameters with hidden latent variables [107].

A SOM is a subclass of Artificial Neural Networks (ANN). SOMs are utilised for dimensionality reduction of large datasets and produce a low-dimensional representation of the input [108]. The SOM network typically implements two nodal layers, namely, the data input layer and a two-dimensional Kohonen layer. The input layer consists of multiple input vectors, which are mapped to the neurons in the Kohonen layer. The association between a neuron and the set of inputs are defined by a weight vector, which determines the spatial location of the neuron in the two-dimensional Kohonen layer space. During training, the weights are updated to change the neuron positions to define well formulated clusters.

A SOM is trained through an unsupervised training process, where the input data are fed to the network through the processing nodes of the input layer. The input vector, $\boldsymbol{X}$, can be described mathematically by the vector set [109]

$$\boldsymbol{X} = [x_i | x_i \epsilon\, \mathbb{R}^2, i = 1, \cdots, M], \tag{3.13}$$

where x_i denotes the i^{th} input signal and M denotes the number of input signals.

The first step in the SOM training process is the initialisation of the neural network weights, w_j, represented mathematically by the expression

$$w_j = \left[w_{ji},\ \ i = 1, 2, 3, \cdots, N\right], \tag{3.14}$$

where w_{ji} denotes the weight associated with the i^{th} input signal and the j^{th} neuron. During the iterative training process, the node weights are adjusted according to the topological relations represented by the input data. With each step a winning neuron is selected using the Euclidean distance, d_j, represented by the expression

$$d_j = \sqrt{\sum_{i=1}^{M}(x_i - w_{ji})^2}. \quad 3.15$$

The winning neuron is the one closest to the randomly selected input vector. The surrounding weights are then updated using the formula

$$\Delta w_{ji} = \eta(t) * T_{j,I(x)}(t) * (x_i - w_{ji}), \quad 3.16$$

where $I(x)$ represents the winning neuron and $\eta(t)$ denotes the learning rate, which determines how fast the weights are updated. The topological neighbourhood function $T_{j,I(t)}$ defines the weight vectors update relationship between the neighbouring neurons and the winning neuron, and is represented mathematically by the function [109]

$$T_{j,I(t)} = exp\left[-\frac{S_{j,I(t)}{}^2}{2\sigma(t)^2}\right], \ \sigma(t) < 1, \quad 3.17$$

where the $S_{j,I(t)}$ represents the lateral Euclidean distance between neurons. Variable σ_t represents a monotonically decreasing scalar function of the time step t.

3.3.2 Optimal number of clusters

Unsupervised partitioning algorithms require a priori specification of the number of clusters. A range of methodologies have been proposed to determine the optimal number of clusters for the various clustering algorithms [110, 111, 112]. However, it is generally acknowledged that the existing metrics for determining the appropriate number of clusters are subjective. Since all partitioning methods require an a priori definition of the number of clusters N^j, it is important to explore methods that can accurately and algorithmically determine this value. Two methods for determining the number of clusters are explored, namely, the elbow point method and silhouette analysis [103, 67, 66].

The widely used heuristic elbow point method [113] provides a graphic representation of the average sum-of-squared distances between intra-cluster samples versus the number of clusters. The method relies on visual interpretation of the benefit derived by increasing the number of clusters. The optimal number of clusters is indicated by an inflection point in the curve. This point denotes the transition from a significant to a marginal improvement in intra-cluster variance.

The silhouette width method determines the distance of separation between resulting clusters. This is typically expressed as a graphical display of the distance between each point within one cluster, to each point between the adjacent clusters. This metric lies in the range [-1, +1], where +1 represents a larger distance between points within adjacent clusters, 0 indicates a close or overlapping proximity and negative values indicate a suboptimal cluster assignment of the observations. The silhouette width algorithm is a function of N^jclusters, which is similar to the elbow point method. However, for each value of N^j, the average silhouette of the observation is calculated. The cluster number with the maximum average silhouette coefficient is then equal to the optimal number of clusters.

The average silhouette width, S, is defined by the mathematical relationship [114]

$$S = \frac{1}{M^i}\sum_{j=1}^{M^j}\sum_{C_{jk}\in C_j}\frac{(b_{C_{jk}}-a_{C_{jk}})}{max(a_{C_{jk}},b_{C_{jk}})}, \quad 3.18$$

where M^i and M^j denote the number of observations and number of clusters, respectively, and $a_{C_{jk}}$ and $b_{C_{jk}}$ denote the mean intra-cluster distance and nearest inter-cluster distance, respectively, for observation C_{jk}.

3.3.3 Cluster validation metrics

The various validation methods that are used to compare the performances of the various clustering algorithms, including the average silhouette width, the Dun index, the Davies-Bouldin criterion and the Calinski-Harabasz index.

The silhouette width is a clustering assignment measure based on the separation and compactness of the clusters. This measure ranges between the interval [-1,1], with 1 denoting perfect clustering results and poorly formed clusters observe a silhouette width near -1 [115]. The silhouette coefficient for the i^{th} element in the dataset for a given a cluster, S_i, is defined by the mathematical relationship [115, 116]

$$S_i = \frac{(b_i-a_i)}{max(a_i,b_i)}, \quad 3.19$$

where a_i and b_i denote the mean intra-cluster distance and nearest inter-cluster distance, respectively, for each i^{th} element.

The Dunn index is a measure based on cluster separation and compactness, which ranges between [0:∞], and should be maximised for optimal results. The Dunn index is represented by the mathematical expression [117]

$$D = \min_{i=1,\dots,N^j}\left\{\min_{j=i+1,\dots,N^j}\left(\frac{diss(c_i,c_j)}{\max_{m=1,\dots,N^j}(diam(c_m))}\right)\right\}, \quad 3.20$$

where

$$diss(c_i,c_j) = \min_{x\in c_i, y\in c_j}\{d(x,y)\}\ and\ diam(c_i) = \max_{x,y\in c_i}\{d(x,y)\}. \quad 3.21$$

N^j denotes the number of clusters, c_i is the i^{th} cluster and $diam(c_m)$ represents the intra-cluster diameter of a cluster m.

The Calinski-Harabasz index is a measure of the average inter-cluster and intra-cluster sum of squares, which should be maximised for optimal cluster assignment [118]. The Calinski-Harabasz index can be described by the mathematical relationship

$$CH = \frac{trace(S_B)}{trace(S_w)}\cdot\frac{n_p-1}{n_p-k}, \quad 3.22$$

where S_B denotes the inter-cluster scatter matric, S_w denotes the intra-cluster scatter matrix and n_p is the number of clusters sampled, with k denoting the number of clusters.

The Davies-Bouldin criterion [119] is a method often used to determine the optimal number of clusters, which is based on an intra-cluster to inter-cluster distance ratio. Since this metric is a measure of cluster compactness, it can also be used to validate the cluster results. The Davies-Bouldin index, DB, is mathematically described as

$$DB = \frac{1}{N}\sum_{i=1}^{N} \underset{j \neq i}{max}\{D_{i,j}\}, \quad 3.23$$

where $D_{i,j}$ is the intra-cluster to inter-cluster distance ratio between the i^{th} and j^{th} clusters. The Davies-Bouldin criterion is optimal for the smallest index value output.

3.4 Wind resource clustering implementation

This section details the clustering implementation process followed to achieve realistic and insightful resource cluster maps for the input of the optimisation framework. In this section, an initial study is executed to determine the ideal resource clustering algorithm to apply for the desired implementation. Once the ideal clustering methodologies have been identified, the clustering process is implemented for the temporal wind power dataset. Thereafter, statistical distributions of the resultant temporal wind power profiles are derived for a predefined set of daily ToU periods. These distributions are translated to ToU feature vectors, which are clustered using the best preforming algorithm.

Figure 3.2 shows an overview of the clustering methodology proposed in this work. The input dataset of mesoscale wind speed profiles is converted to power profiles, which are normalised relative to a wind turbine rating of 3.3 MW, with the single turbine power curves are detailed in Table 2.2.

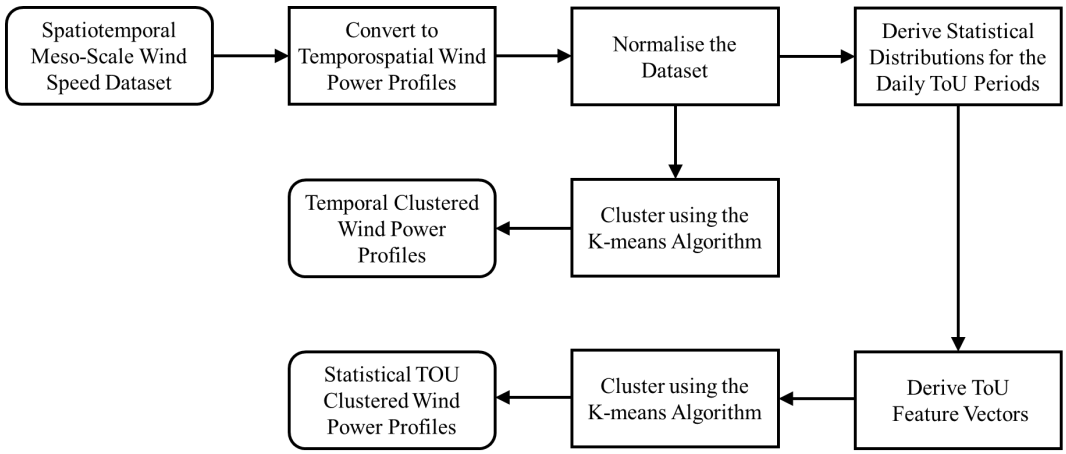


Figure 3.2: Overview of the proposed clustering methodology.

3.4.1 Wind resource cluster methodology performance evaluation

This section explores the implementation of various clustering algorithms for two targeted input approaches for the clustering of wind power resource datasets. The various clustering methods are explored and compared using four main evaluation criteria, namely, execution time, the Calinski-Harabasz criterion, the Davies-Bouldin criterion and the silhouette width index. Table 3.1 compares the various clustering methodologies based on the four evaluation criteria for a temporal resource input dataset. The k-means algorithm shows a superior execution time. The SOM approach shows superior results according to both the Davies-Bouldin criterion and the silhouette width index, however, the difference between the k-means, PAM and SOM results are marginal.

Table 3.2 compares the various clustering methodologies based on the four evaluation criteria, for a ToU feature-based resource input dataset. For this approach, the k-means algorithm shows superior results for all performance metrics. Therefore, to ensure uniformity across the clustering implementation, the k-means algorithm is selected for both the temporal and statistical methods.

Table 3.1: Comparison of the performance evaluation metrics calculated for the temporal cluster approach.

Algorithm	Execution time [seconds]	Calinski-Harabasz	Davies-Bouldin	Silhouette width
K-means	75.500	527.724	1.712	0.3565
PAM	983.850	547.030	1.733	0.3563
CLARA	223.240	515.457	1.912	0.3252
Agglomerative	13152.010	316.276	1.951	0.2984
Diana	6252.030	316.276	1.951	0.2984
SOM	26986.860	543.029	1.685	0.3842

Table 3.2: Comparison of the performance evaluation metrics calculated for the ToU feature-based approach.

Algorithm	Execution time [seconds]	Calinski-Harabasz	Davies-Bouldin	Silhouette width
K-means	0.098	2779.761	1.098	0.4505
Model	0.498	1202.049	2.682	0.0765
PAM	7.907	2776.699	1.149	0.4175
CLARA	0.230	2594.237	1.204	0.3998
Agglomerative	5.200	2402.371	1.224	0.3794
Diana	162.250	2303.231	1.243	0.3725
SOM	4.347	2650.274	1.276	0.3866

3.4.2 Cluster formations obtained with temporal wind power profiles

In this section, the clustered results are presented for the implementation of the k-means algorithm, using the squared Euclidean distance metric, for the temporal wind power input dataset.

To implement the clustering process, a priori selection of the appropriate number of clusters for the classification of the eleven REDZs is required. A novel hybrid elbow point and silhouette width method is implemented to identify the optimal number of clusters. Figure 3.3 depicts the initial number of clusters range, chosen between the 2 and 20. The elbow point method is implemented for this initial range, resulting in a graphical inflection point from which a second, smaller range is determined. This is depicted in Figure 3.3, where a piecewise linear function is superimposed on the elbow line-plot to determine the reduced range of cluster numbers. The graph suggests an optimal number of clusters in the range between 10 and 12. The average silhouette widths for each number within this range is computed to determine the optimal number of clusters. These average silhouette widths are depicted in Figure 3.4(a), which suggests that the appropriate choice for the number of clusters is ten. This is supported by the silhouette widths shown in Figure 3.4(b) for the individual cluster members. Few cluster members depict negative silhouette widths, which is indicative of optimal classification.

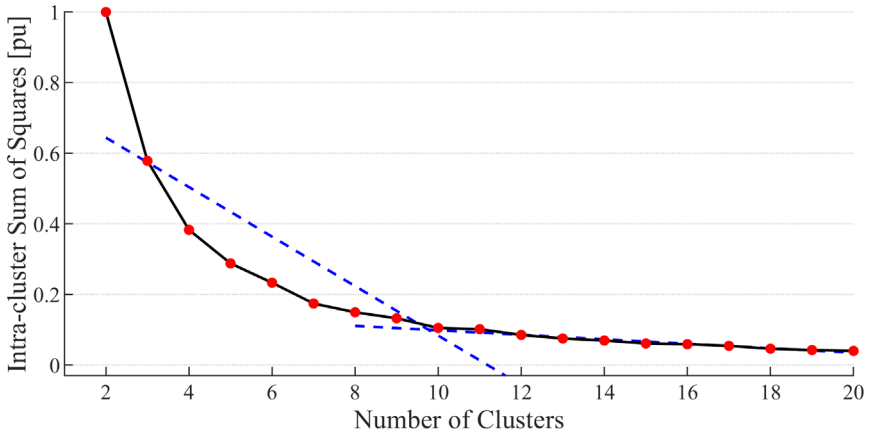
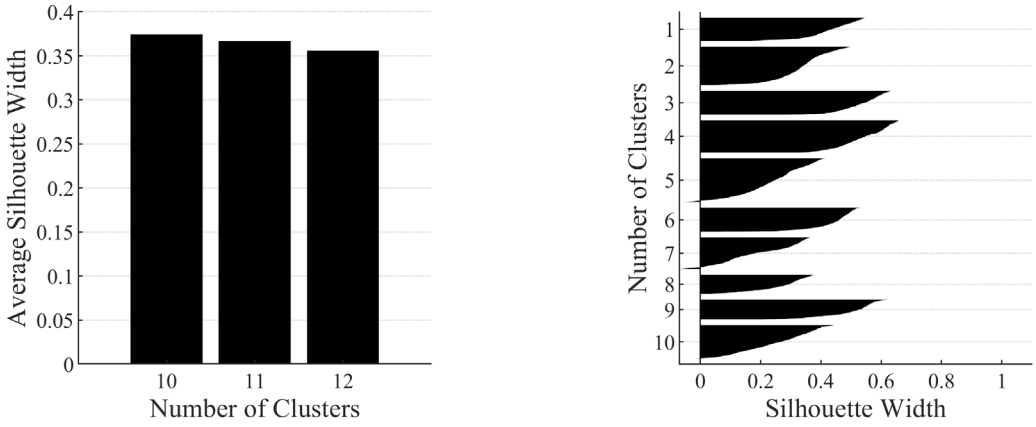


Figure 3.3: Elbow plot of the total intra-cluster sum of squares versus the number of clusters for clustering the temporal wind power profiles of all REDZs using the k-means algorithm.



(a) (b)

Figure 3.4: Silhouette widths for clustering the temporal profiles of all REDZs using the k-means algorithm: (a) Average silhouette width versus the number of clusters for the cluster range suggested by the elbow point method, (b) Individual silhouette widths for 10 clusters.

Figure 3.5 shows the cluster distribution obtained by clustering the temporal wind power profiles of all REDZs into 10 clusters using the k-means algorithm. Overall, the clusters align with the geographic regions represented by the individual REDZs. However, the cluster formations associated with the individual REDZs exhibit a low degree of dispersion and granularity. Most REDZs reflect a highly homogenous character in the sense that the associated cluster formations consist of a single or low number of different clusters. The cluster map for the Springbok REDZs for instance, shown in Figure 3.6, depicts a single cluster ranging from a coastal to inland region, despite a wide range in elevation levels.

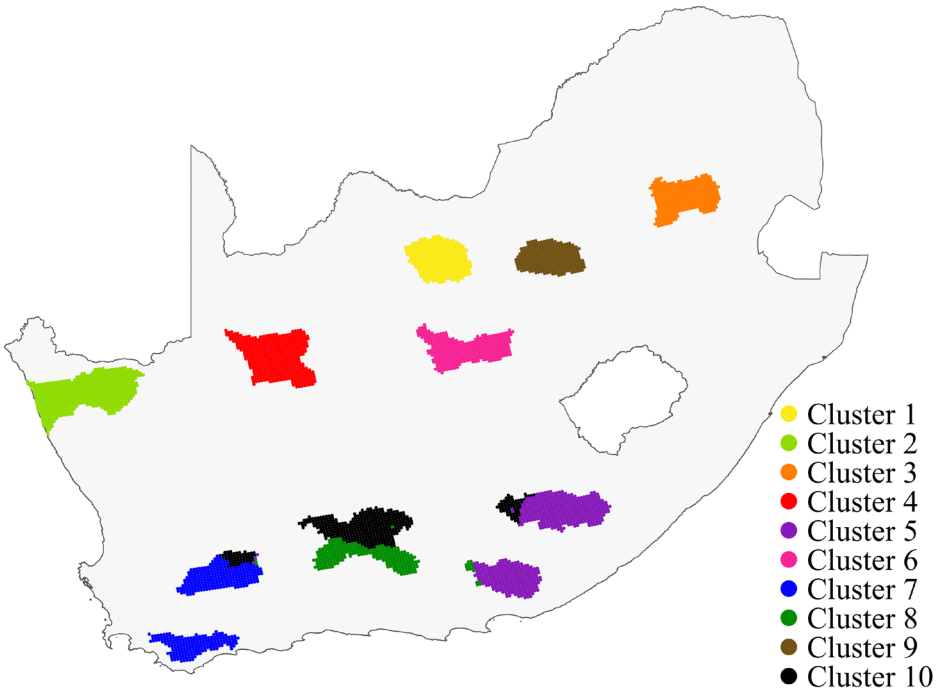


Figure 3.5: Cluster distribution results obtained by clustering the temporal wind power profiles over all REDZs.

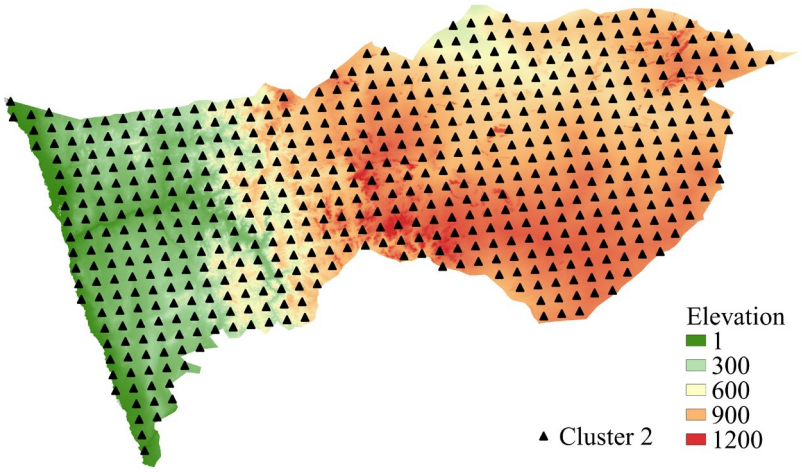


Figure 3.6: Cluster distribution obtained by clustering the temporal wind power profiles using the k-means algorithm, overlayed onto the elevation map of the Springbok REDZs.

3.4.3 Cluster formations obtained with the statistical wind power feature vectors

This section describes the design, implementation and performance evaluation of a wind resource classification methodology that aligns with operational power balance in the context of a ToU framework. The key research question addressed in this section focuses on comparing the performance of classical spatiotemporal clustering [58, 61, 64] with a novel feature-based clustering approach. In this approach, the features are defined in terms of the statistical properties of the spatiotemporal wind power profiles for a given set of daily and seasonal ToU periods. In this context, it is recognised that the resource characteristics and grid impact of wind generation profiles can be interpreted with reference to the daily and seasonal cycles exhibited by the demand profiles, and that some ToU periods are more important than others.

The results are presented for the wind resource associated with the South African Renewable Energy Development Zones, using a mesoscale spatiotemporal wind atlas dataset as the resource input. The results obtained by clustering the ToU feature vectors, using the k-means algorithm, are compared with results delivered by clustering the temporal power profiles. More particularly, the geographical cluster distributions obtained with the two methodologies are compared by interpreting the statistical properties of the clusters in the context of temporal grid support.

The translation of the temporal power profiles to ToU feature vectors reduces the dimensionality of the spatiotemporal dataset significantly, i.e., from 175 296 wind power observations per spatial coordinate to a total of 12 feature values per spatial coordinate. These features consist of the standard deviation σ and mean μ of the temporal wind power profiles for the peak, standard and off-peak periods for the high and low demand periods, summarised in Table 2.3.

Figure 3.7 shows a density scatterplot of the μ, σ feature pairs defined in Table 2.3 for all of the normalised temporal wind power profiles. The extracted statistical parameters vary over a considerable range, such that μ varies from 0.0733 to 0.7754 and σ varies from 0.1660 to 0.4585, respectively. This spread is indicative of a diverse range of wind resource characteristics for the different ToU periods, which may be leveraged in siting optimisations for grid support objectives.

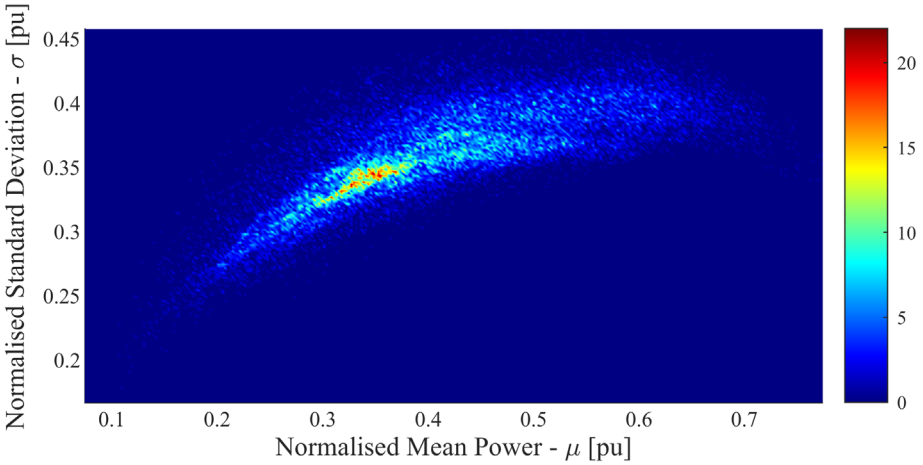


Figure 3.7: Density scatterplot of the mean and standard deviation of the normalised wind power profiles for the daily ToU periods. The dark blue regions reflect the lowest densities, while the red regions reflect the highest densities.

Figure 3.8 depicts the initial number of clusters range, namely, between the 2 and 20. The elbow point method is implemented for clustering the ToU feature vectors, with results suggesting an optimal number of clusters in the range between 10 and 12. The average silhouette widths shown in Figure 3.9(a) for this range suggests that 10 clusters is an appropriate choice for the optimal number of clusters. This is supported by the individual silhouette widths shown in Figure 3.9(b), where 8 of the 10 clusters comprise of a minor percentage of members with negative silhouette widths.

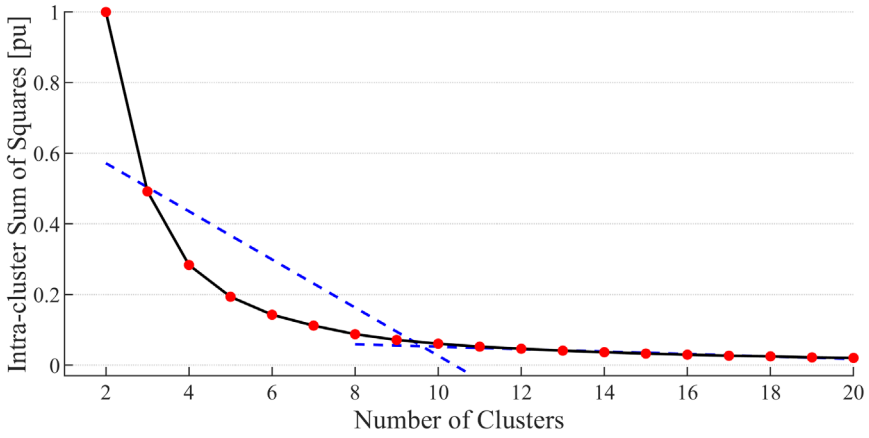
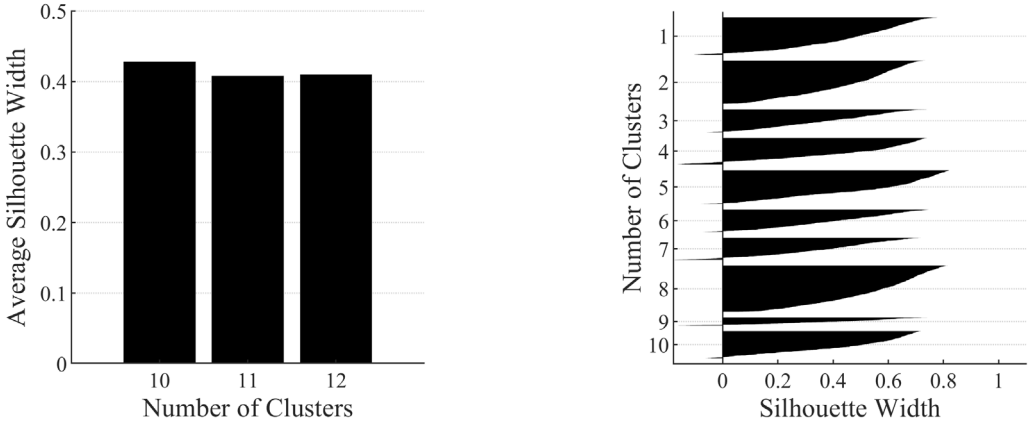


Figure 3.8: Elbow plot of the total intra-cluster sum of squares versus the number of clusters for clustering the ToU feature vectors of all REDZs using the k-means algorithm.



(a) (b)

Figure 3.9: Silhouette widths for clustering the ToU feature vectors of all REDZs using the k-means algorithm: (a) Average silhouette width versus the number of clusters for the cluster range suggested by the elbow point method, (b) Individual silhouette widths for 10 clusters.

Figure 3.10 depicts the results obtained by clustering the ToU feature vectors for all REDZs into 10 clusters using the k-means algorithm. The cluster formations associated with the individual REDZs exhibit a much higher degree of dispersion and granularity compared to the result shown in Figure 3.5 for the temporal profiles. The cluster distribution shown in Figure 3.11 for the Springbok REDZ, for instance, includes 10 different clusters, while this region is represented by only 1 cluster using the temporal clustering approach, shown in Figure 3.6.

Overall, it is clear that the two methodologies deliver distinctly different results, especially in terms of the granularity of the resultant cluster formations. This may be of importance in the use of clustering as a resource classifier in studies aimed at the optimal geospatial allocation of wind generation capacity.

Figure 3.11 provides geographic depictions of all eleven REDZs clustered and overlayed onto the wind speed resource map for the South African region, depicted in Figure 2.2. The graphic images depict strong correlation patterns between the ToU feature based clustering results and the underlying wind profile patterns. This depicts a significant result, as the ToU feature based approach significantly decreases the dimensionality of the input resource dataset as well as increases the correlation between

the allocated clusters and the underlying geographic and climatic features. The formulated clusters also define regions which increase alignment with the ToU energy profiles, which will greatly increase the accuracy of the final optimisation results to support RE plant placement for grid support objectives.

Figure 3.13 depicts the resultant average daily wind power profiles associated with the (a) annual, (b) high and (c) low demand seasons over the five-year temporal span for the temporal wind power cluster input profiles. Figure 3.14 depicts the resultant average daily wind power profiles associated with the (a) annual, (b) high and (c) low demand seasons over the five-year temporal span for the ToU feature-based cluster input profiles. The profiles are averaged over the 5-year period and depict distinct variations in the output results when comparing both methodologies. The averaged high and low demand season graphs are plotted over the peak, standard and off peak ToU periods.

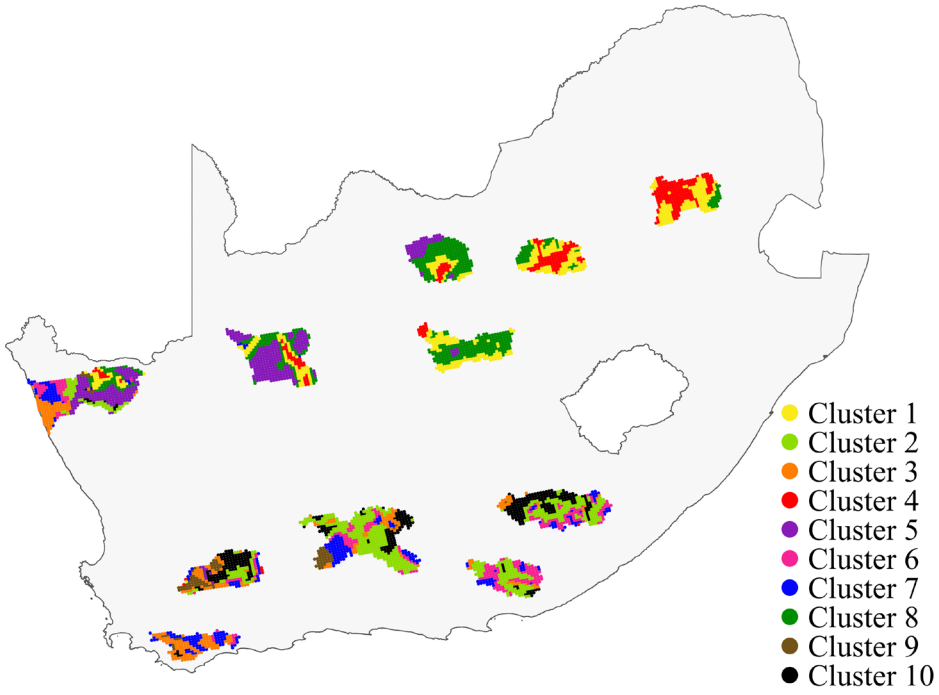


Figure 3.10: Cluster distribution results obtained by clustering the ToU feature vectors over all REDZs.

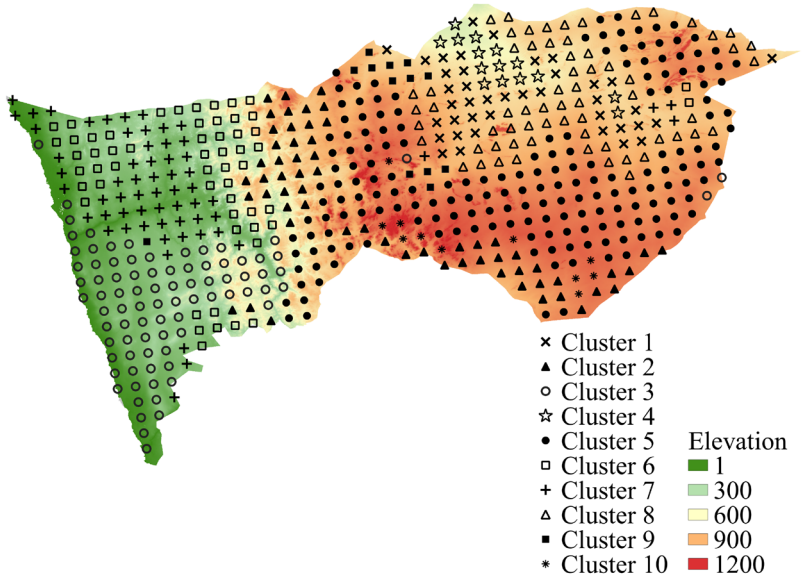
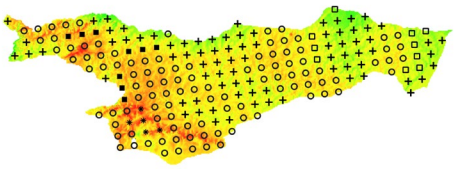
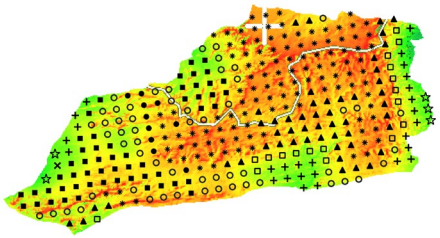


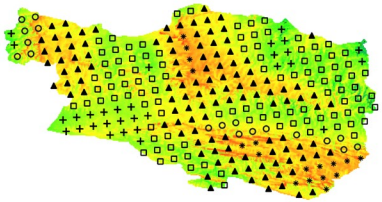
Figure 3.11: Cluster distribution obtained by clustering the ToU feature vectors of a zoomed in elevation map for the Springbok REDZs using the k-means algorithm.



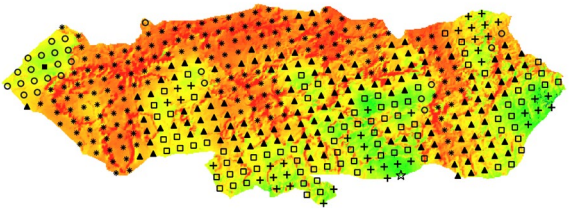
Overberg REDZ 1



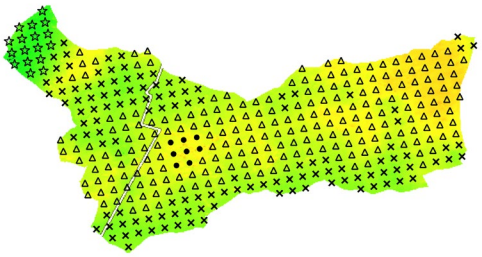
Komsberg REDZ 2



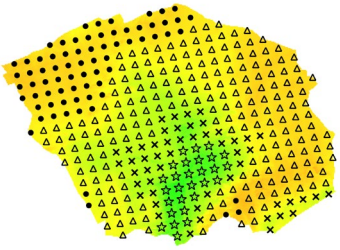
Cookhouse REDZ 3



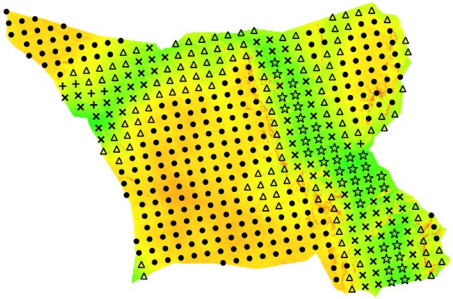
Stormberg REDZ 4



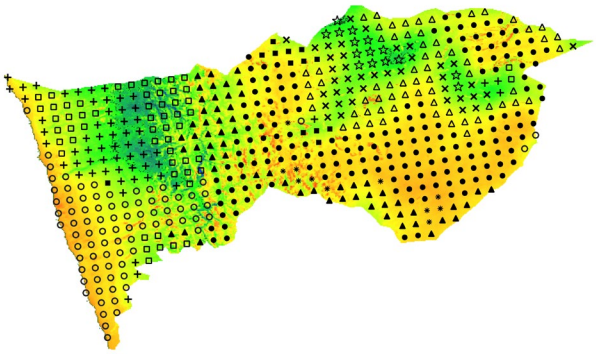
Kimberly REDZ 5



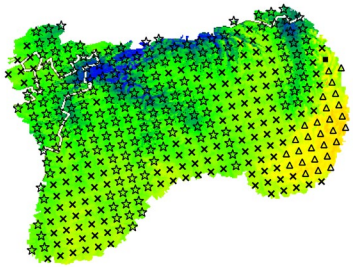
Vryburg REDZ 6



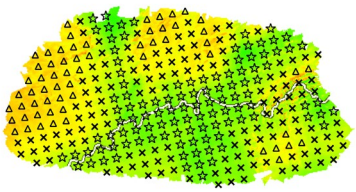
Upington REDZ 7



Springbok REDZ 8



Emalaheni REDZ 9



Klerksdorp REDZ 10

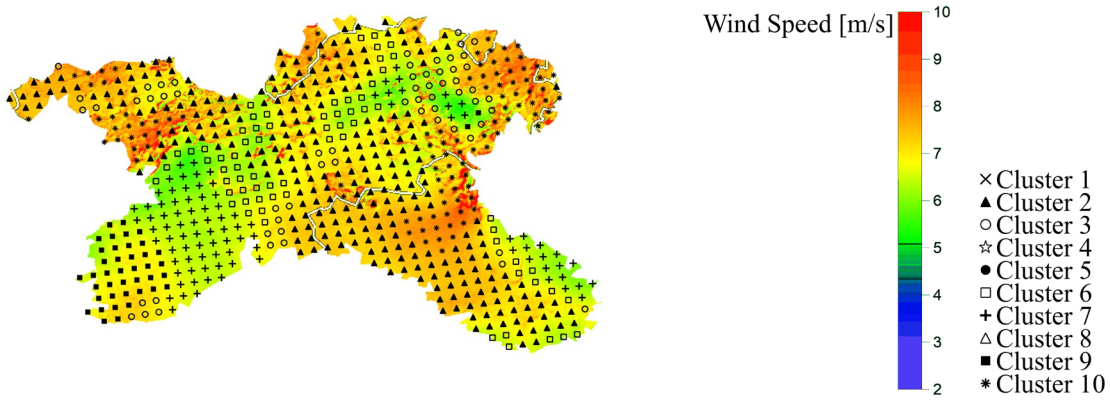
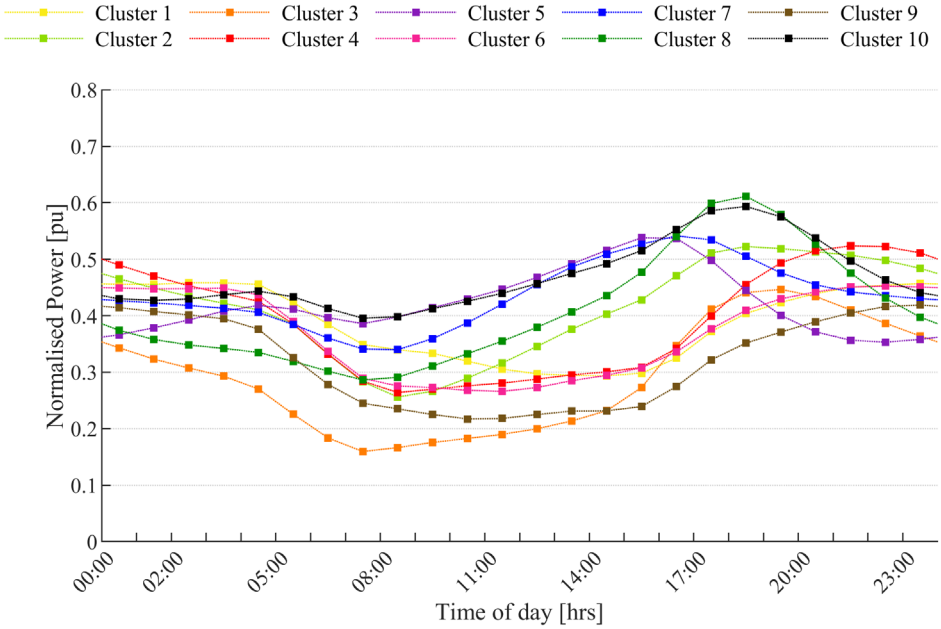
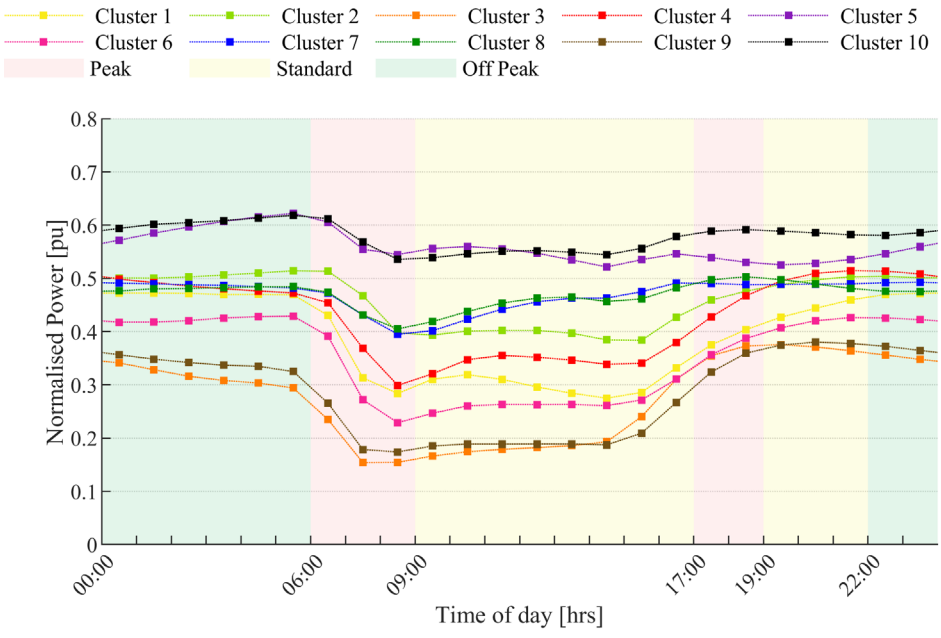


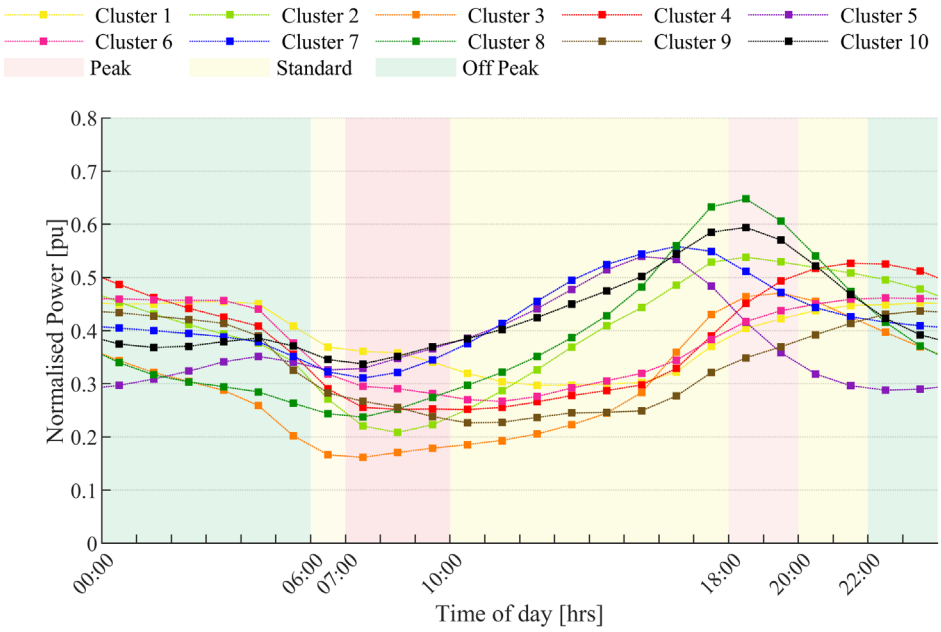
Figure 3.12: The clustered profiles plotted as an overlay onto the average wind speed plot for South Africa based on the ToU input vectors.



(a) Annual period

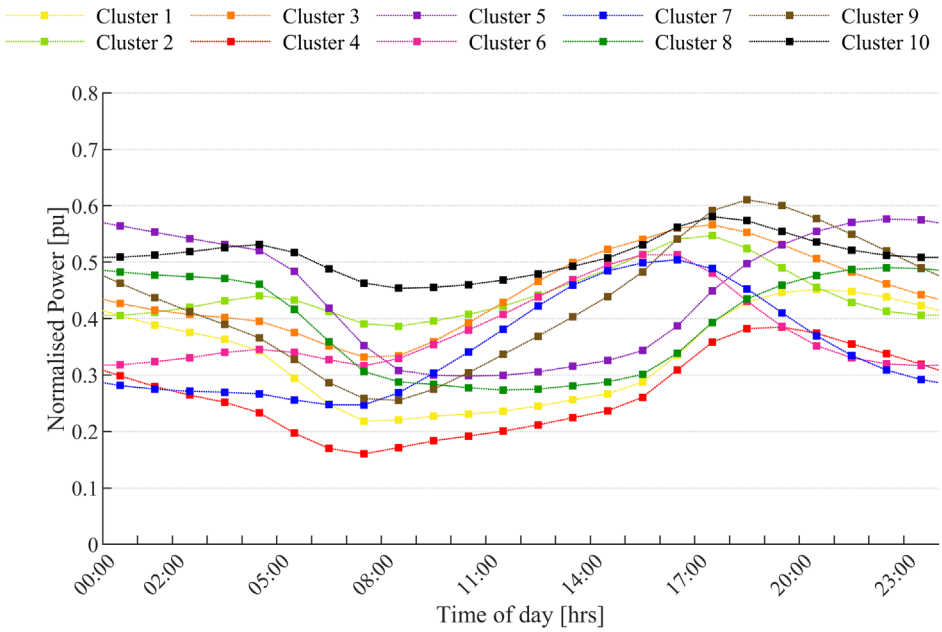


(b) High Demand Season

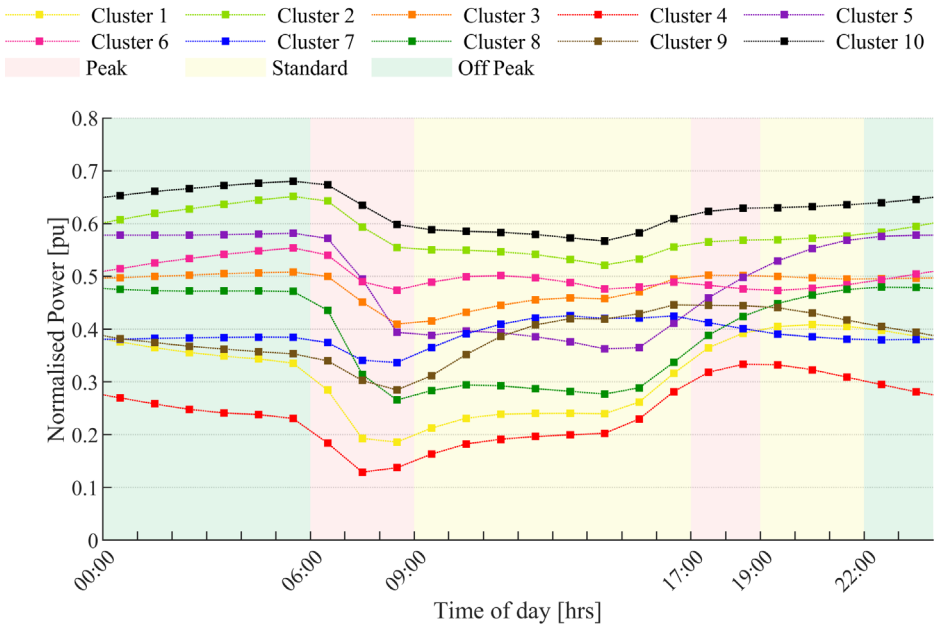


(c) Low Demand season

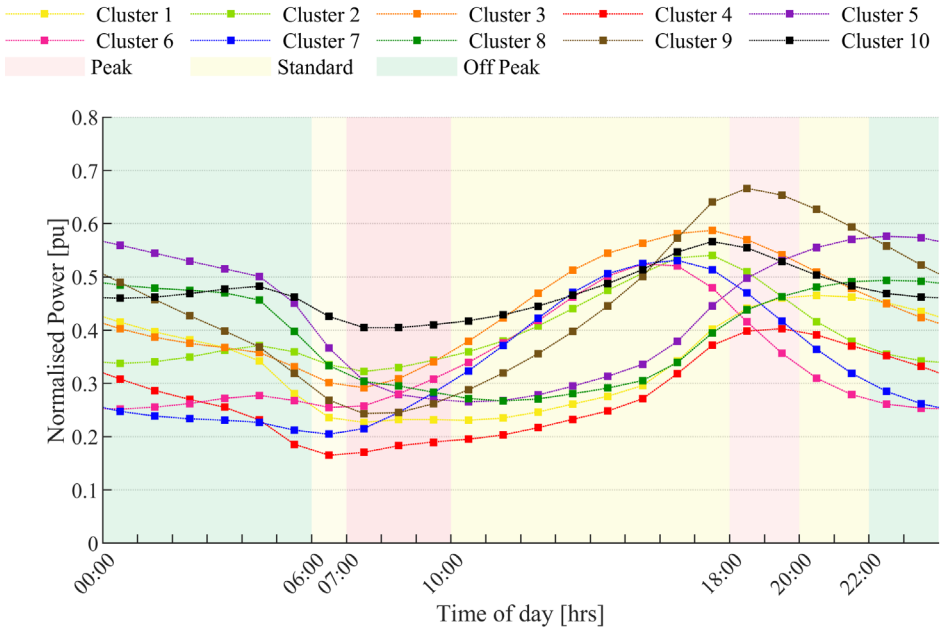
Figure 3.13: The clustered daily wind power profiles depicted for the (a) annual, (b) high and (c) low demand seasons using the temporal clustering approach.



(a) Annual Period



(b) High Demand Season



(c) Low Demand Season

Figure 3.14: The clustered daily wind power profiles depicted for (a) annual, (b) high and (c) low demand seasons using the ToU feature-based clustering approach.

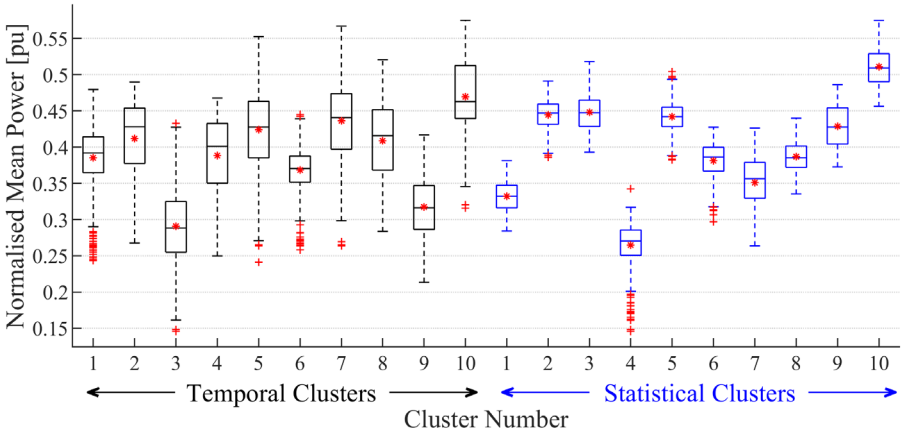
3.4.4 Statistical comparison and analysis

Statistical comparisons of the cluster formations are performed for the temporospatial and ToU feature-based approaches. This is achieved by comparing the means of the wind power profiles associated with the individual clusters using standard boxplots. In the granularity of the plots, the overall mean wind power of each cluster is indicated by a red dot within each boxplot, the median of each mean cluster is

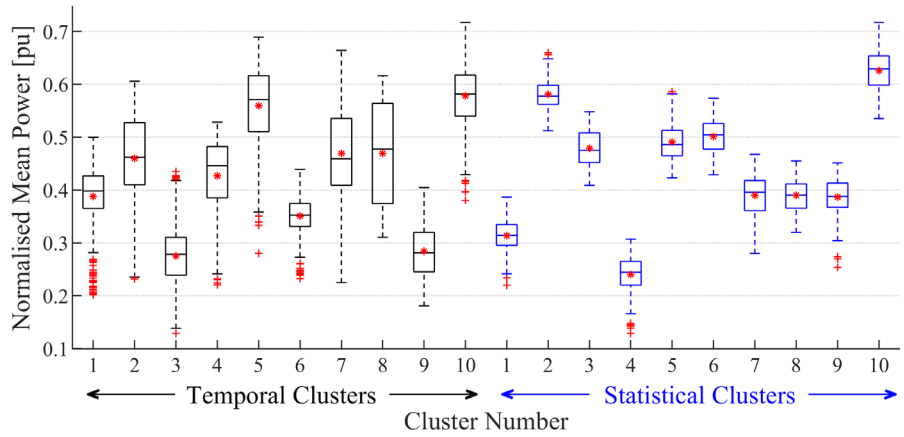
indicated by the centre horizontal line, and the bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. The whiskers extend to the extreme data points, excluding the outliers, which are indicated by the tail symbols depicted outside the whisker ranges. The results are presented for the wind power profiles of the full five-year temporal span of the input dataset.

Figure 3.15(a) depicts boxplots of the means of the wind power profiles associated with the individual clusters, averaged over the five annual periods. On average, the clusters obtained with the ToU feature vectors have lower intra-cluster variance of the mean power of each cluster member compared to the clusters obtained with the temporal profiles. Figure 3.15(b) and Figure 3.15(c) show boxplots of the means of the wind power profiles associated with the individual clusters, averaged over the high demand season and low demand seasons, respectively. On average, for both demand seasons, the clusters obtained with the ToU feature vectors have lower intra-cluster variance of the mean power for each cluster member compared to the clusters obtained with the temporal profiles. The high demand season exhibits higher mean power values compared to the annual period and low demand season for both clustering approaches. However, the mean wind power of the profiles associated with the individual clusters vary considerably more for the high demand season compared to the low demand season.

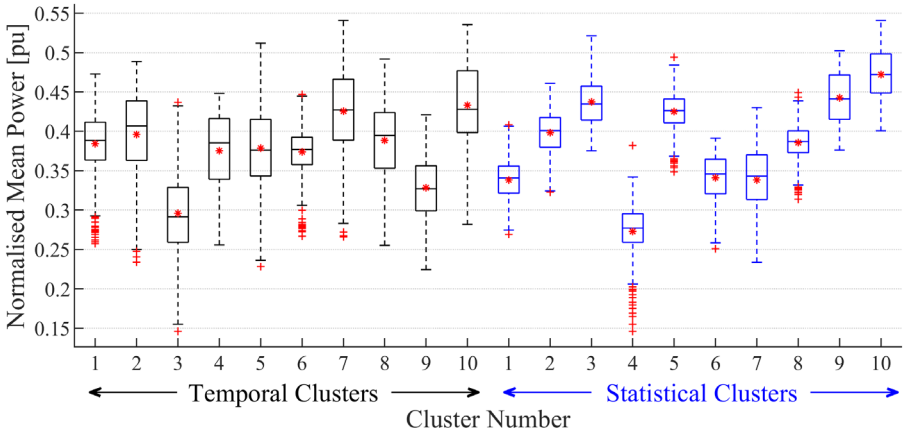
Table 3.3 compares the mean wind power of each cluster for the average annual, high demand season and low demand season periods. The ToU feature-based approach delivers the cluster with the highest overall means for all cases, namely, cluster 10.



(a) Annual period



(b) High demand season



(c) Low demand season

Figure 3.15: Boxplots of the normalised mean power of the profiles associated with the individual clusters obtained by clustering the temporal profiles and ToU feature vectors of all REDZs, for the (a) annual, (b) high demand season and (c) low demand season averaging intervals.

Table 3.3: Mean wind power of the clusters obtained by clustering the temporal profiles and ToU feature vectors of all REDZs, for the annual, high demand season and low demand season averaging intervals.

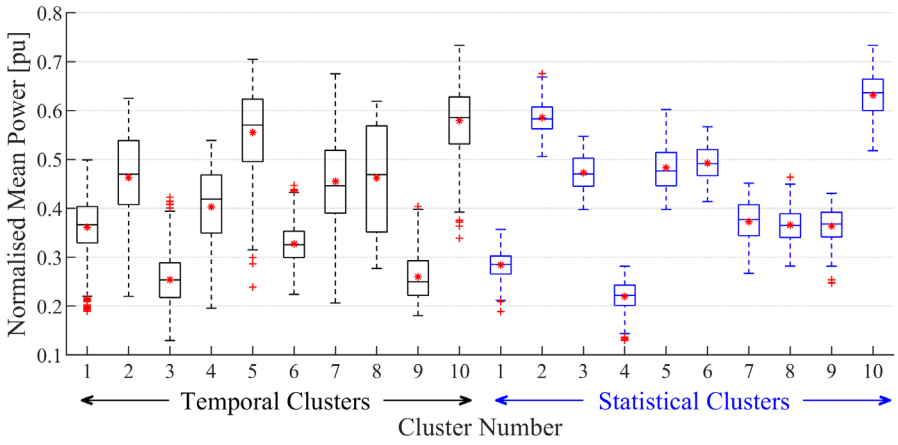
Cluster Number	Temporal			Statistical		
	Annual	HDS	LDS	Annual	HDS	LDS
1	0.385	0.388	0.384	0.332	0.314	0.338
2	0.412	0.460	0.396	0.444	0.581	0.399
3	0.291	0.275	0.296	0.448	0.479	0.437
4	0.389	0.427	0.376	0.265	0.241	0.273
5	0.424	0.559	0.379	0.442	0.491	0.425
6	0.368	0.351	0.374	0.382	0.501	0.341
7	0.437	0.469	0.426	0.351	0.390	0.338
8	0.409	0.469	0.389	0.387	0.390	0.386
9	0.317	0.284	0.328	0.428	0.386	0.443
10	0.470	0.578	0.433	**0.511**	**0.626**	**0.472**

Figure 3.16 depicts the boxplots of the means of the wind power profiles associated with the individual clusters for the (a) peak, (b) standard and (c) off peak ToU periods, averaged over the high demand season. For each ToU period, when comparing the temporal vs ToU feature based approaches, the ToU feature based method produces better overall cluster results with each cluster formation showing lower variations in the mean wind power profiles as well as higher overall mean wind power profiles for each cluster. Figure 3.17 depicts the boxplots of the means of the wind power profiles associated with the individual clusters for the (a) peak, (b) standard and (c) off peak ToU periods, averaged over the low demand season.

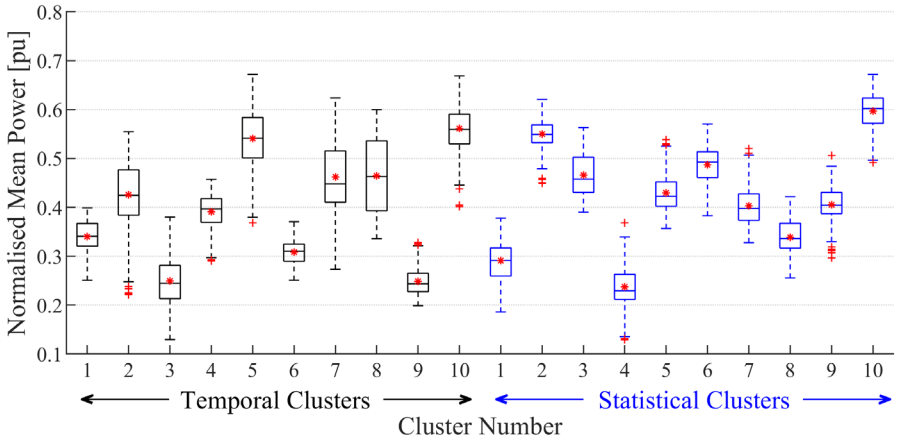
Overall, the results presented in Figure 3.16 and Figure 3.17 show that the clusters can assist in identifying appropriate geographic regions to optimise energy yield and variability targets for specific demand seasons and ToU periods. The clusters obtained with the ToU feature vectors generally rate superior compared to the clusters obtained with temporal profiles, especially in the sense that the means

of profiles associated with the individual clusters have lower variance. Clustering by ToU periods using the ToU feature vectors effectively partitions locations into clusters with distinctive seasonal wind characteristics. Cluster 10 obtained with the ToU feature vectors, as shown in Figure 3.16 for instance, has the highest normalised mean power for all averaging periods. The ToU feature vectors also capture the cluster with the lowest normalised mean power over the two demand seasons, namely, cluster 4 shown in Figure 3.16 for all ToU periods in the high demand season. Table 3.4 compares the mean wind power of the clusters for the peak ToU period for the high demand season and low demand season averaging periods. The ToU feature vectors deliver the clusters with the highest overall means for both cases, namely, cluster 10.

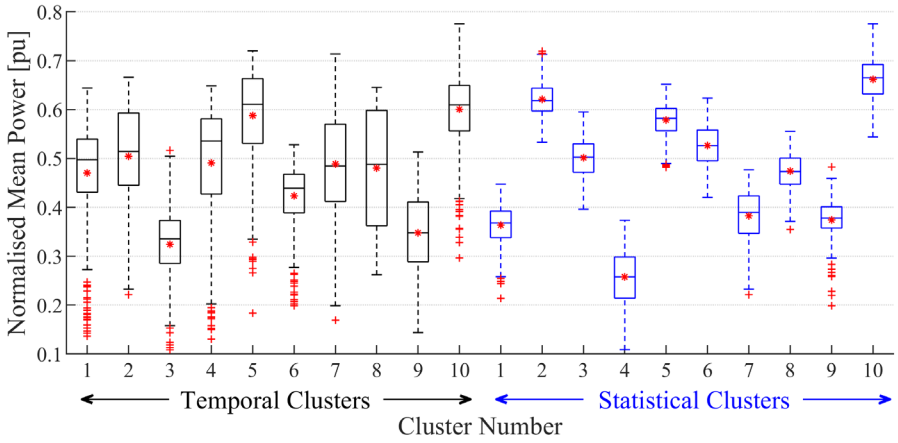
The derived cluster maps have application in the optimal spatial allocation of wind generation in the context of grid support. It is evident from the results that clustering based on ToU statistical features presents an opportunity to leverage the temporal diversity of spatiotemporal wind resources.



(a) Peak Periods

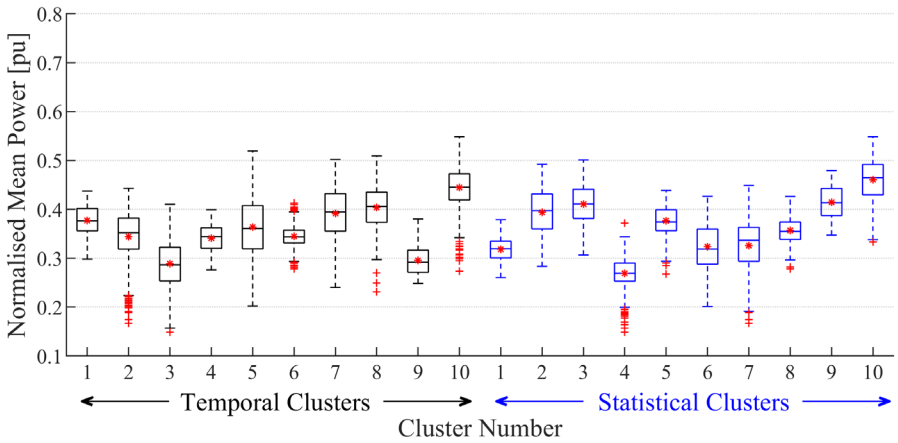


(b) Standard Periods

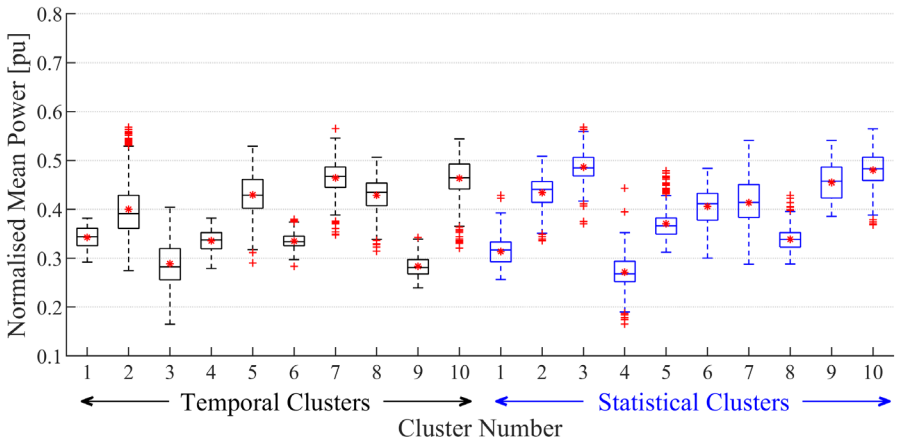


(a) Off peak Periods

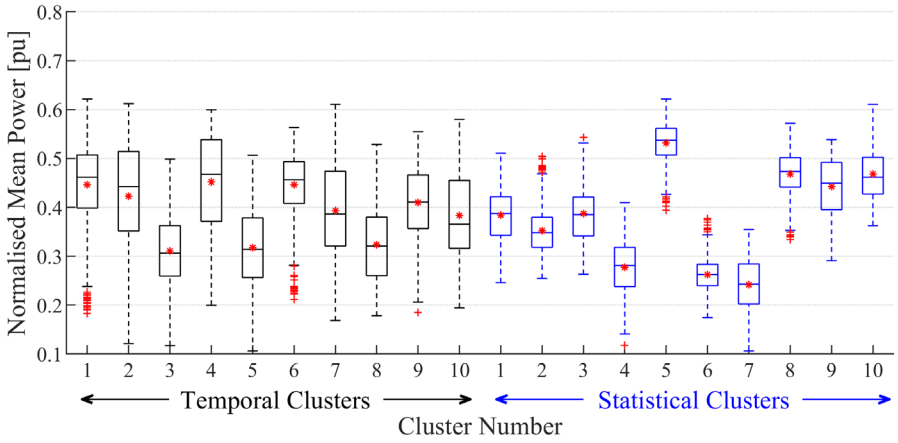
Figure 3.16: High demand season Boxplots of the normalised mean power in the (a) peak, (b) standard and (c) off peak periods of the profiles associated with the individual clusters obtained by clustering the temporal profiles and ToU feature vectors of all REDZs.



(a) Peak Periods



(b) Standard Periods



(c) Off Peak Periods

Figure 3.17: Low demand season Boxplots of the normalised mean power in the (a) peak, (b) standard and (c) off peak periods of the profiles associated with the individual clusters obtained by clustering the temporal profiles and ToU feature vectors of all REDZs.

Table 3.4: Mean wind power over the peak ToU period of the clusters obtained by clustering the temporal profiles and ToU feature vectors of all REDZs, for the high demand and low demand season averaging intervals.

Cluster Number	Temporal		Statistical	
	HDS	LDS	HDS	LDS
1	0.361	0.377	0.284	0.318
2	0.463	0.344	0.585	0.394
3	0.254	0.289	0.473	0.411
4	0.403	0.341	0.220	0.269
5	0.555	0.364	0.484	0.377
6	0.327	0.345	0.493	0.324
7	0.455	0.392	0.373	0.326
8	0.462	0.404	0.366	0.357
9	0.260	0.296	0.364	0.414
10	0.579	0.445	**0.632**	**0.461**

Table 3.5 summarises the performance of the ToU feature based approach and the temporal input approach across various metrics. Comparison of the input dataset sizes shows a significant reduction in the ToU feature input.

Table 3.5: Comparison of the cluster evaluation metrics for the ToU features and the temporal wind power profiles inputs.

Input dataset	Input dataset dimensions	Execution time [seconds]	Calinski-Harabasz	Davies-Bouldin	Silhouette width
ToU features	12 x 4470	0.0977	2779.761	1.098	0.451
Temporal wind power	175 296 x 4470	101.160	527.724	1.712	0.356

Since the ToU feature based approach achieved significantly superior results, this approach is chosen as the optimal clustering methodology to be used for this research application.

3.4.5 Wind power clustering conclusion

Characterising and classifying spatiotemporal wind resources by clustering of the temporal wind power profiles awards equal weights to the contribution of the inter- and intra- cluster distance metrics for all hours. This inadvertently places a high importance on the long off-peak periods during the late evening and early morning, compared to the relatively shorter standard and peak ToU periods in the daily demand profile. The contribution of the relatively long low demand season to these distance metrics, similarly, outweighs the contribution for the shorter high demand season. However, from a system operations perspective the energy yield and variance of the cumulative wind power generation profile are interpreted with reference to the demand profile. Demand profiles are variable, and exhibit daily and seasonal characteristics, such as peak ToU periods where the operational risks are typically most severe.

This section explores the translation of spatiotemporal wind power profiles into feature vectors defined in terms of the statistical properties of these profiles for a predefined set of ToU periods. The proposed approach allows for spatiotemporal wind resource data to be translated to spatial cluster formations, with due cognisance of the temporal characteristics of the demand profile. The resulting clustered resource maps are of value for the siting of wind generation capacity such that the power generation profiles deliver optimal grid support in the context of the daily and seasonal residual load profile. Furthermore, for the case study of the spatiotemporal wind resource dataset associated with the South African REDZs, the translation to feature vectors reduces the dimensionality of the dataset to be clustered significantly, i.e., from 175 296 15-minute wind power observations for each of the 4470 spatial coordinates to 12 ToU feature values per spatial coordinate.

The cluster distributions obtained with the ToU feature vectors are compared with distributions delivered by clustering the spatiotemporal profiles using the k-means algorithm. This delivers a significant result in the sense that the distributions obtained with the ToU feature vectors across the eleven geographical REDZs, for the same number of clusters, show significantly higher granularity and diversity. It is concluded that the clustering of ToU feature vectors, compared to the clustering of raw temporal profiles, distinguishes more finely between the daily and seasonal temporal properties of neighbouring geographic locations. Overall, the results show that the power profiles associated with the individual clusters obtained by the feature vector approach have similar or higher means and similar or lower standard deviations compared to the clusters obtained with spatiotemporal profiles.

The ToU based clustering approach has potential as a data reduction pre-processing step in optimising the site locations of future wind energy plants for optimal grid support in the context of the diurnal and seasonal characteristics of the aggregated load profile. The ToU periods used in the case study can be readily amended to represent the operating characteristics of different systems, or incorporate different criteria, such as time-dependent transmission constraints, real-time pricing considerations, etc.

3.5 Solar resource clustering implementation

In this section a similar clustering implementation process is followed to achieve realistic and insightful solar resource cluster maps for the input of the optimisation framework. Table 3.6 and Table 3.7 show the results of a performance evaluation study completed for various clustering algorithms. In this section, the clustering process is implemented for a temporal solar power dataset. Thereafter, statistical distributions of the resultant temporal solar power profiles are derived for a predefined set of daily ToU

periods. These distributions are translated to ToU feature vectors, which are clustered using the best performing algorithm.

Figure 3.18 shows an overview of the solar clustering methodology proposed in this work.

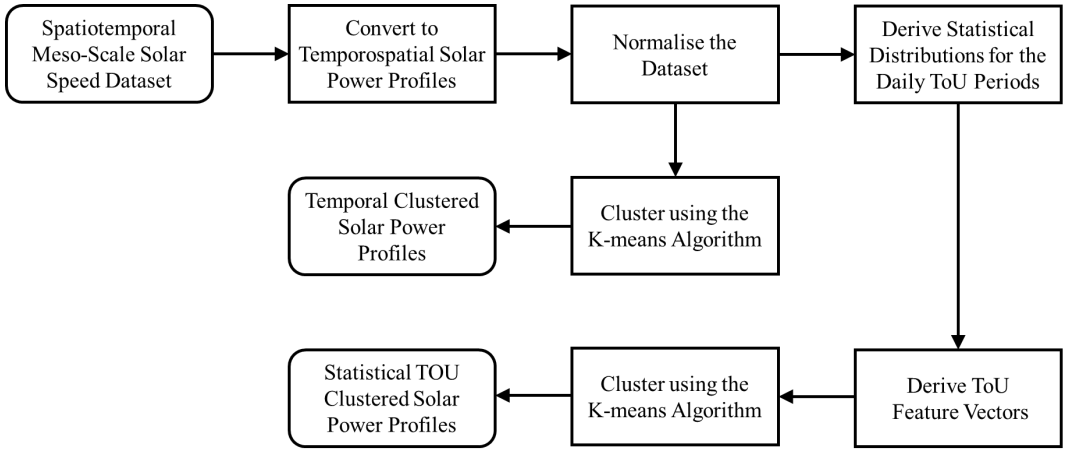


Figure 3.18: Overview of the proposed solar clustering methodology.

3.5.1 Solar resource cluster methodology performance evaluation

This section explores the implementation of various clustering algorithms for two targeted input approaches for the clustering of solar power resource datasets. The various clustering methods are explored and compared using four main evaluation criteria. Table 3.6 and Table 3.7 summarise the results, with the main evaluation criterion including the algorithm execution time, the Calinski-Harabasz criterion, the Davies-Bouldin index, the silhouette width index. Table 3.6 compares the results for the instance where the temporal solar power profiles are used as the input variables. Table 3.7 compares the results for the instance where the 13 ToU feature profiles are used as input, summarised in Table 2.4.

The k-means algorithm shows a superior execution time for both instances. The PAM approach shows superior results according to the silhouette width index. However, the difference between the PAM, CLARA and Agglomerative Hierarchical algorithm results are marginal.

Table 3.6 compares the various clustering methodologies based on the four evaluation criteria, for a ToU feature-based resource input dataset. For this approach, the PAM algorithm shows superior results for the silhouette width index. Agglomerative Hierarchical clustering performs well according to the Davies-Bouldin and silhouette width metrics, although the results are close to that of PAM and k-means algorithms for the silhouette width and Calinski-Harabasz. The PAM method is also similar in value to the Agglomerative Hierarchical method for the Davies-Bouldin criteria. The Agglomerative Hierarchical clustering method has a very long execution time when compared with the shortest execution time, which is for the k-means algorithm.

Table 3.7 shows that the k-means algorithm achieves superior results over all performance metrics. To ensure uniformity across the clustering implementation, the k-means algorithm is selected for both the temporal and statistical methods.

Table 3.6: Comparison of the cluster evaluation metrics for the temporal solar power profiles inputs.

Algorithm	Execution time [seconds]	Calinski-Harabasz	Davies-Bouldin	Silhouette width
K-means	8.955	763.640	1.511	0.4314
PAM	190.759	772.253	1.469	0.4396
CLARA	48.554	739.207	1.476	0.4170
Diana	7998.720	611.437	1.473	0.4070
Agglomerative	6455.510	772.345	1.469	0.4397
Fuzzy C	298.001	1051.724	2.036	0.3155
SOM	7663.004	726.057	1.493	0.4115

Table 3.7: Comparison of the cluster evaluation metrics for the solar power 13 ToU features profile inputs.

Algorithm	Execution time [seconds]	Calinski-Harabasz	Davies-Bouldin	Silhouette width
K-means	0.030	4484.118	0.679	0.6928
Model	0.570	2172.934	1.735	0.4410
PAM	4.835	4519.905	0.849	0.6766
CLARA	0.100	4413.404	0.867	0.6684
Diana	144.780	2451.240	0.956	0.5369
Agglomerative	2.000	4295.248	0.832	0.6487
Fuzzy C	0.594	4297.365	1.007	0.5904
SOM	5.440	4552.645	0.836	0.6826

3.5.2 Cluster formations obtained with temporal solar power profiles

In this section, the results are presented for the implementation of the k-means algorithm, using the squared Euclidean distance metric, for the temporal solar power input dataset. The dataset comprises of 4470 spatial locations, with each location representing a five-year temporal span, namely, from 2009 to 2013, with an hourly temporal resolution.

To implement the clustering process, a priori selection of the appropriate number of clusters for the classification of the eleven REDZs is required. The hybrid elbow point and silhouette width method is again implemented in order to identify the optimal number of solar clusters. In the final all-encompassing model, the ideal number of solar clusters is also determined as a function of the final optimisation objectives.

Figure 3.19 depicts the initial range of the number of clusters, which is selected between the cluster range 2 and 20. The elbow point method is implemented for the initial range, resulting in a graphical inflection point, from which a second smaller range for the number of clusters is determined. This is depicted in Figure 3.19, where piecewise linear functions are superimposed on the elbow plot in order to determine the reduced range of cluster numbers. The graph suggests an optimal number of clusters in the range between 8 and 10. The average silhouette widths for each number within this range is computed to inform the choice of the optimal number of clusters. These average silhouette widths are depicted in Figure 3.20(a), which suggests that the appropriate choice for the number of clusters is 10. This is supported by the silhouette widths shown in Figure 3.20(b) for the individual cluster members. Few cluster members depict negative silhouette widths, which is indicative of optimal classification.

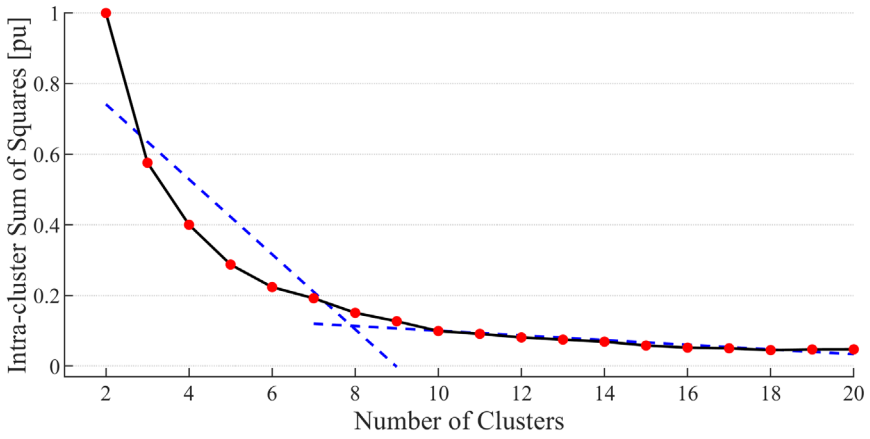
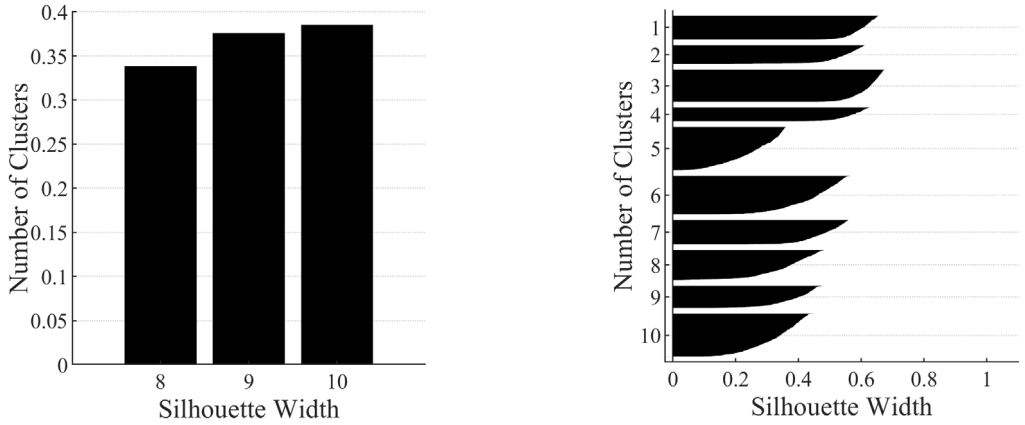


Figure 3.19: Elbow plot of the total intra-cluster sum of squares versus the number of clusters for clustering the temporal profiles of all REDZs using the k-means algorithm.



(a) (b)

Figure 3.20: Silhouette widths for clustering the temporal profiles of all REDZs using the k-means algorithm: (a) Average silhouette width versus the number of clusters for the cluster range suggested by the elbow point method, (b) Individual silhouette widths for 10 clusters.

Figure 3.21 shows the cluster distribution obtained by clustering the temporal solar power profiles of all REDZs into 10 clusters using the k-means algorithm. Overall, the clusters align with the geographic regions represented by the individual REDZs. The cluster formations associated with the individual REDZs exhibit a low degree of dispersion and granularity. Most REDZs reflect a highly homogenous character in the sense that the associated cluster formations lie within a single REDZs.

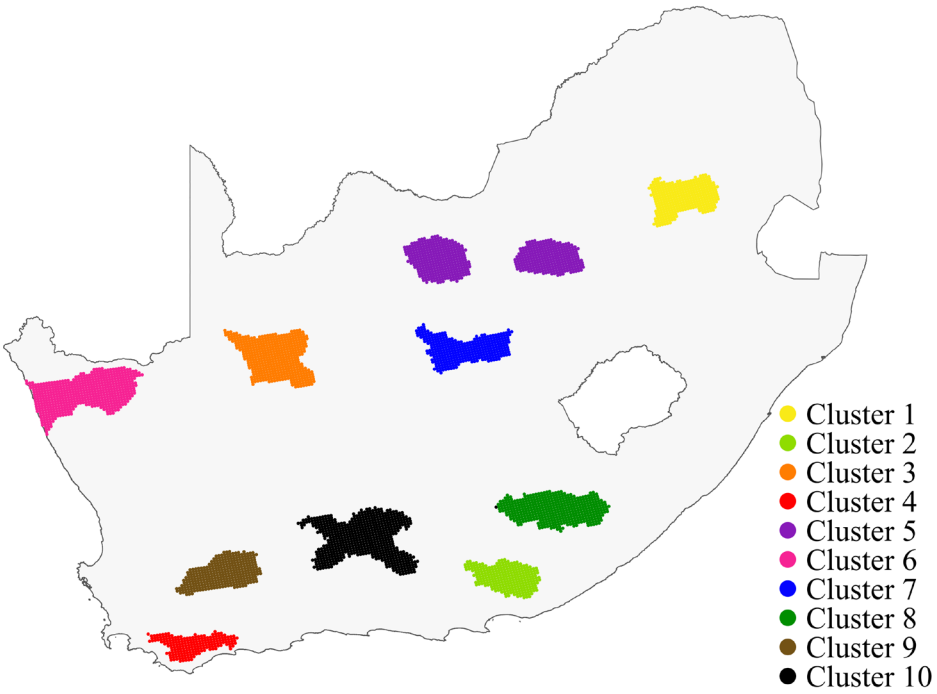


Figure 3.21: Cluster distribution obtained by clustering the temporal solar power profiles of all REDZs into 10 clusters.

3.5.3 Cluster formations obtained with the statistical solar power feature vectors

This section describes the design, implementation and performance evaluation of a solar resource classification methodology that aligns with operational power balance in the context of a ToU framework. In this approach, the features are defined in terms of the statistical properties of the spatiotemporal solar power profiles for a given set of daily and seasonal ToU periods. In this context, it is recognised that the resource characteristics and grid impact of solar generation profiles can also be interpreted with reference to the daily and seasonal cycles exhibited by the demand profiles, and that some ToU periods are more important than others.

The results are presented for the solar resource power profiles associated with the South African REDZs, using a mesoscale spatiotemporal solar power resource dataset as the input. The results obtained by clustering the ToU feature vectors using the k-means algorithm are compared with results delivered by clustering the temporal power profiles using the k-means algorithm. More particularly, the geographical cluster distributions obtained with the two methodologies are compared by interpreting the statistical properties of the clusters in the context of temporal grid support.

The translation of the temporal power profiles to ToU feature vectors reduces the dimensionality of the spatiotemporal dataset significantly, i.e., from 43 824 hourly solar power observations per spatial coordinate to a total of 26 feature values per spatial coordinate. These features consist of the standard deviation σ and mean μ of the temporal solar power profiles for the peak, standard and off-peak periods, during the morning, midday and evening periods for the high and low demand periods, summarised in Table 2.4. The daily ToU feature-based profiles derived for the solar resource are divided into the morning, midday and evening periods because the solar bell curve shows important movement patterns throughout the morning, midday and evening periods dependent on the time of year.

Figure 3.22 depicts the initial number of clusters range, namely, between the 2 and 20. The elbow point method is implemented for clustering the ToU feature vectors, with results suggesting an optimal number

of clusters in the range between 10 and 12. The average silhouette widths shown in Figure 3.23(a) for this range suggests that 10 clusters is an appropriate choice for the optimal number of clusters. This is supported by the individual silhouette widths shown in Figure 3.23(b), where 8 of the 10 clusters comprise of a minor percentage of members with negative silhouette widths.

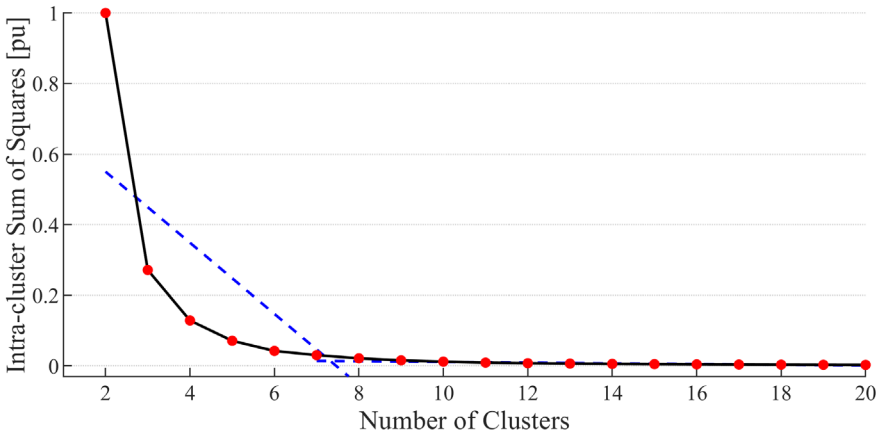
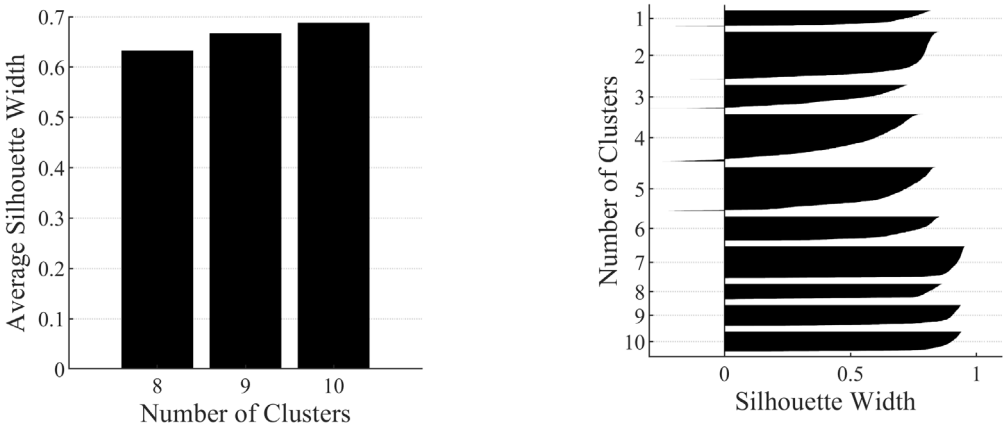


Figure 3.22: Elbow plot of the total intra-cluster sum of squares versus the number of clusters for clustering the ToU feature vectors of all REDZs using the k-means algorithm.



(a) (b)

Figure 3.23: Silhouette widths for clustering the ToU feature vectors of all REDZs using the k-means algorithm: (a) Average silhouette width versus the number of clusters for the cluster range suggested by the elbow point method, (b) Individual silhouette widths for 10 clusters.

Figure 3.24 shows the cluster distribution obtained by clustering the ToU feature vectors of all REDZs into 10 clusters using the k-means algorithm. The cluster formations associated with the individual REDZs exhibit a slightly increased degree of dispersion and granularity compared to the result shown in Figure 3.21 for the temporal profiles.

Figure 3.25 provides geographic depictions of all eleven REDZs clustered using the ToU approach and overlayed onto the solar GHI resource map for the South African region, depicted in Figure 2.3. The graphic images depict correlation patterns between the ToU feature based clustering results and the underlying solar profile patterns. The solar resource profile is vastly different to that of the wind resource

profile, whereby the solar production curve is centred around the midday point. Due to this, the solar profiles display less variability compared to the wind power profiles and cluster variations are mainly determined by the movement of the solar bell curve. It is observed that the clustered profiles for both methods depict low levels of granularity, which is accounted to the nature of the geographically tied resource. The solar GHI does not change as sporadically as the wind speed profiles in various regions. However, the clustered ToU feature-based implementation does show merit in the large dataset reduction as well as capturing finer gradient resolution change over the change in geographic GHI as shown in Figure 2.3. This can be seen more clearly in Figure 3.25 for Springbok and Beaufort West, REDZ 8 and 11, respectively. Figure 3.25, for both REDZs, shows clear changes in the underlying GHI intensity and this translates to different cluster assignemnts in these areas. This shows that the ToU clustering algorithm idetifies small changes within the PV profile and assigns the clusters accordingly. For example, Figure 3.25(a) depicts a definite drop in solar GHI in one specific area and this area is awarded a different cluster assignment, namely, cluster 8. The fluctuation changes in GHI could be due to changes in elevation, slope, roughness, terain or general surface topology.

In Figure 3.25, specific zones are highlighted based on the areas which depict a change in the average solar GHI per REDZs, such as, Springbok and Stormberg. This does not apply for the temporally clustered profiles, depicted in Figure 3.21, since all zones comprise of only one clustered profile, despite the change in solar GHI within REDZs, which is depicted in Figure 2.3.

The Springbok REDZ comprises of 1 cluster for the temporally clustered case, namely, cluster 6. The Springbok REDZ predominately comprises of two clusters for the ToU feature-based approach, namely, clusters 1 and 9. Comparing these averaged profiles per cluster for both clustering methods, cluster 6, shown in Figure 3.26(a), depicts a highest mean power for the daily temporal profile for the averaged cluster. The cluster 9 for the ToU feature-based approach, shown in Figure 3.27(a), also depict the highest mean values, however cluster 9 depicts a higher mean and a higher standard deviation than cluster 1. Cluster 1, shown in Figure 3.27(a), depicts a lower mean with a lower standard deviation compared to cluster 9. This shows that the ToU feature based approach was able to capture finer granularity within this REDZs, resulting in clusters which better represent the zones solar GHI profile.

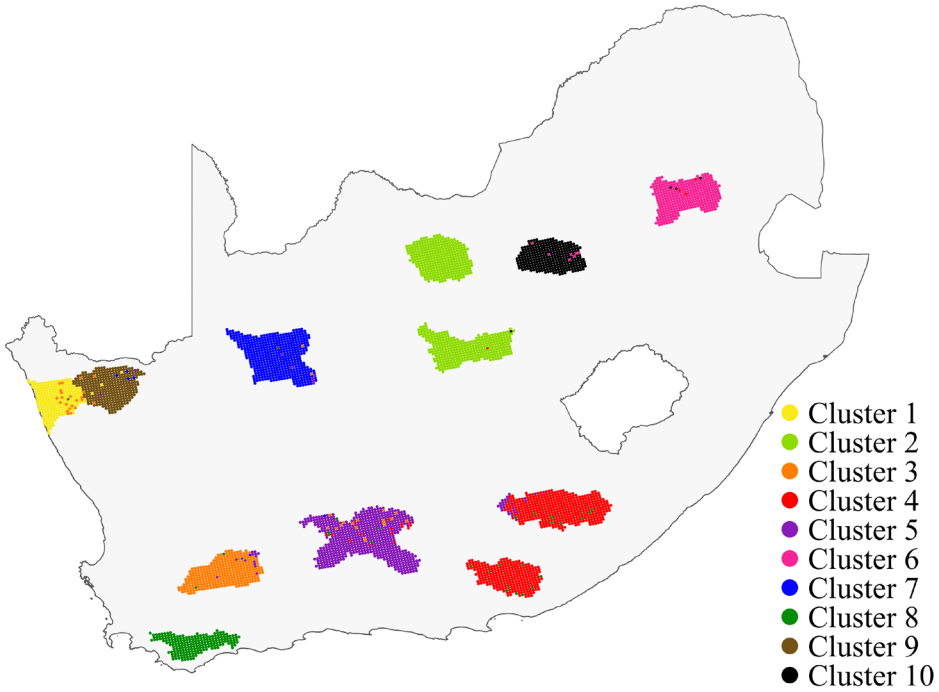
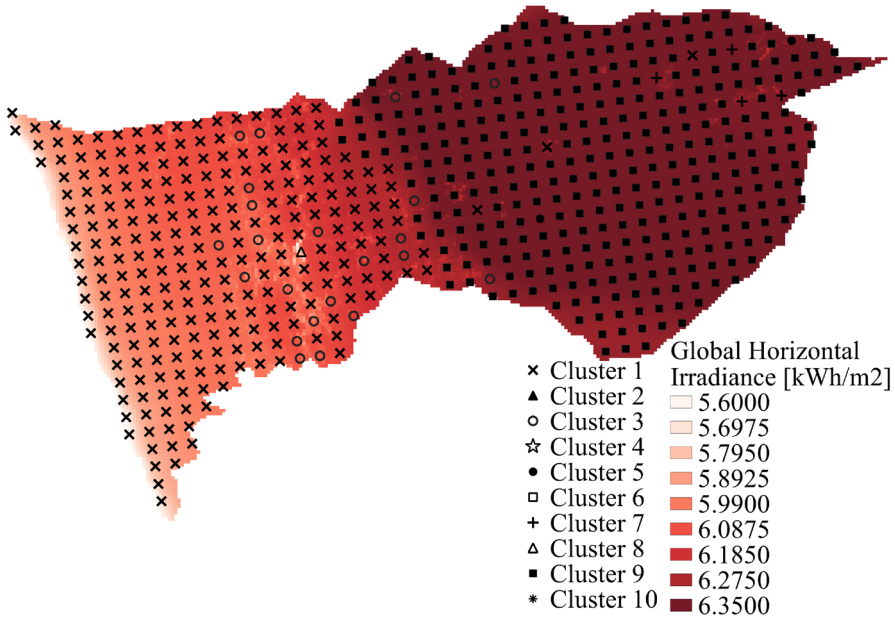
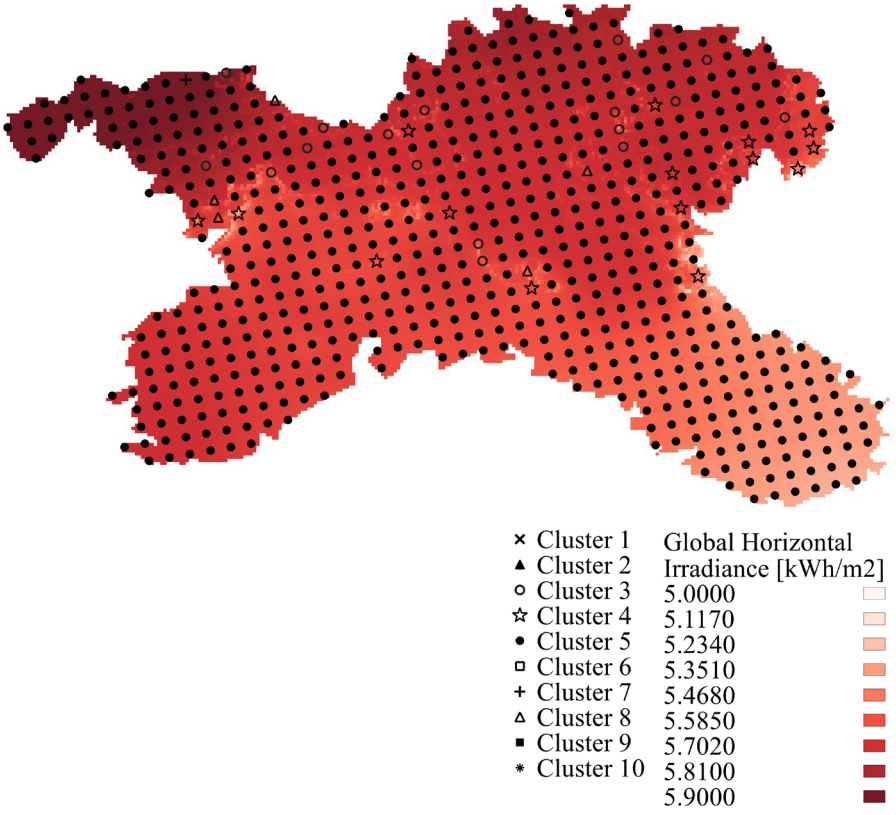


Figure 3.24: Cluster distribution obtained by clustering the ToU feature vectors of all REDZs into 10 clusters.

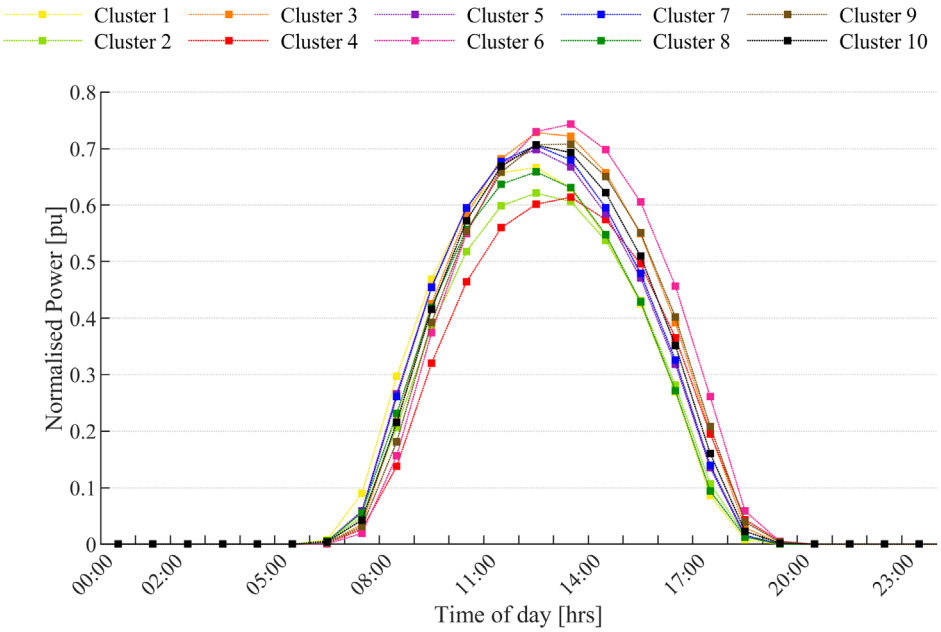


(a) Springbok REDZ 8

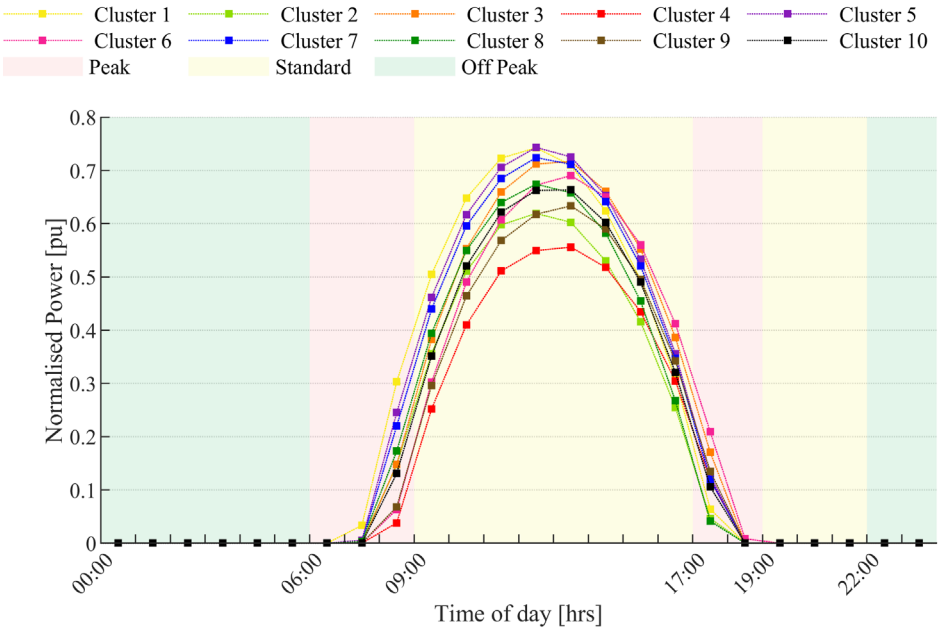


(b) Beaufort West REDZ 11

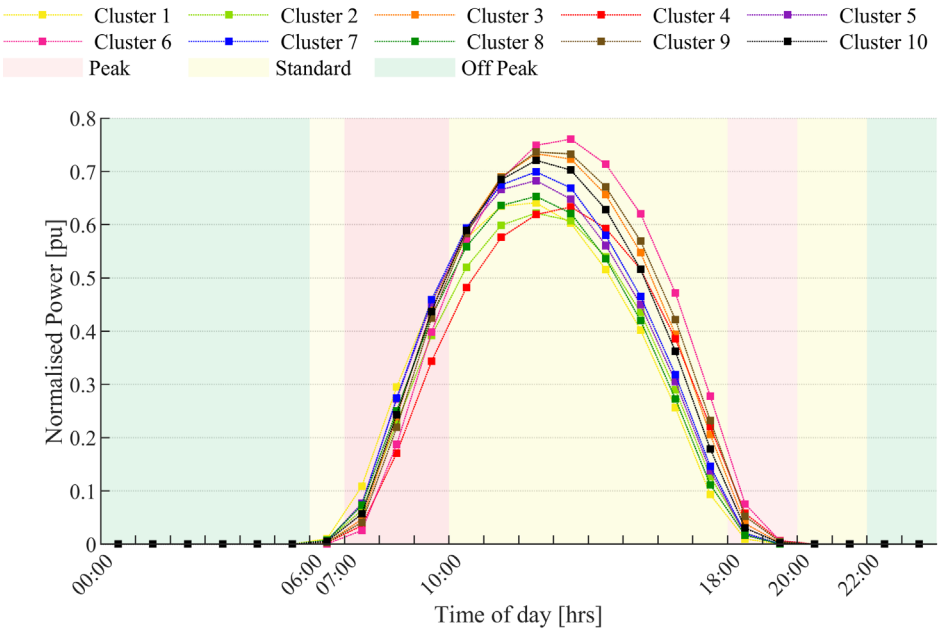
Figure 3.25: The clustered profiles plotted as an overlay onto the global horizontal irradiation image for South Africa based on the ToU input vectors for REDZs 8 and 11.



(a) Annual Period

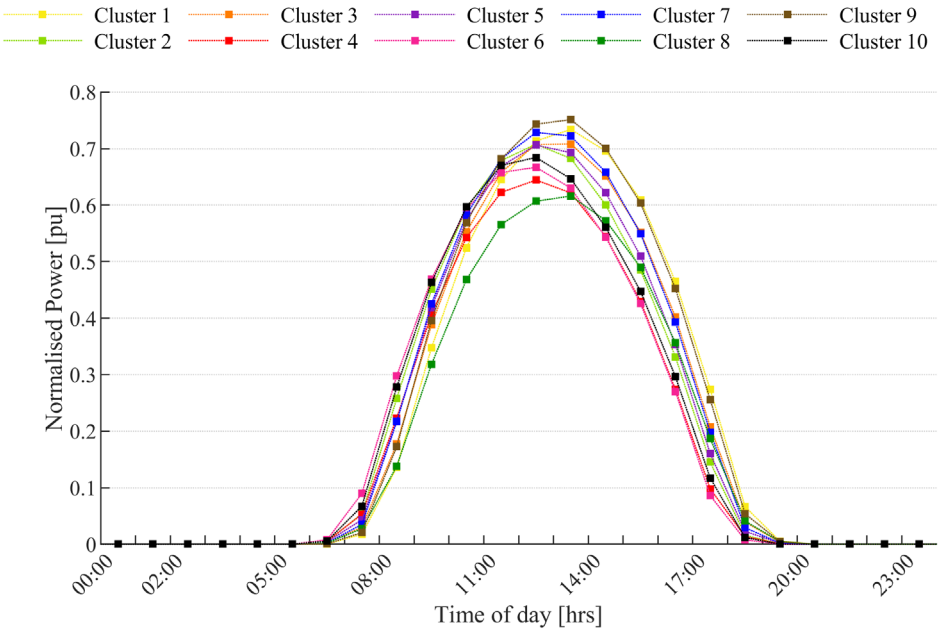


(b) High Demand Season

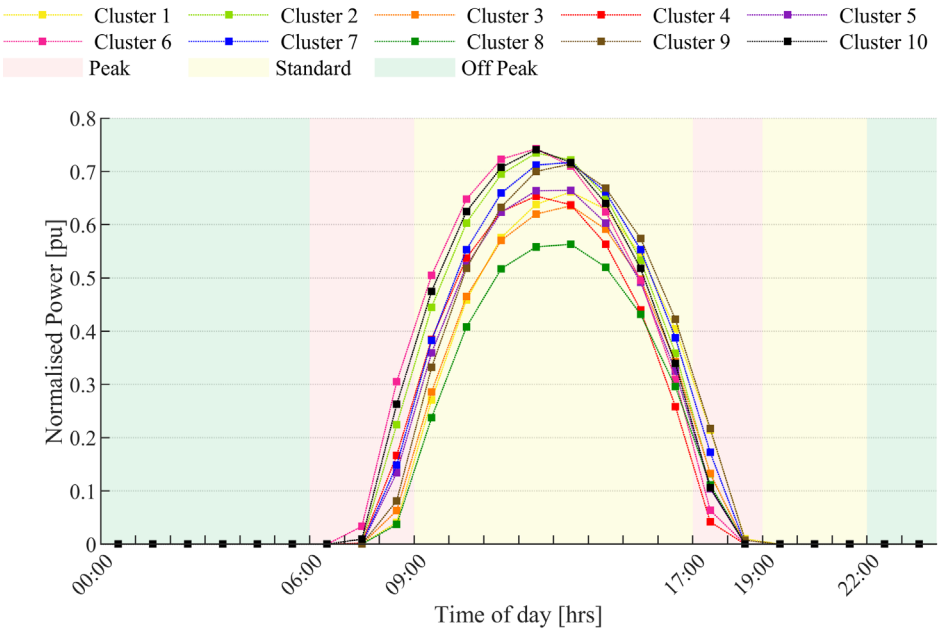


(c) Low Demand Season

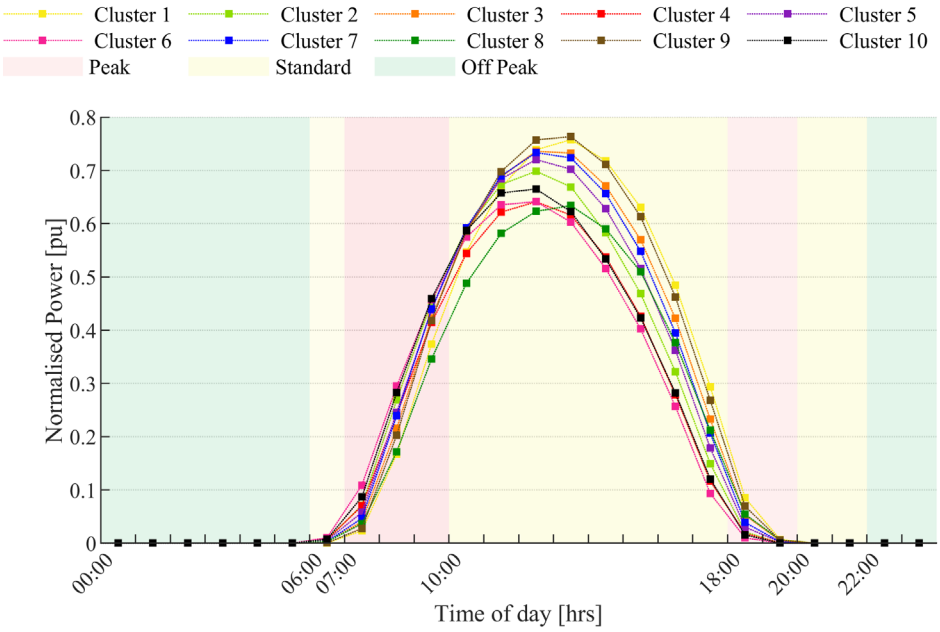
Figure 3.26: The clustered daily solar power profiles depicted for the (a) annual, (b) high and (c) low demand seasons using the temporal clustering approach.



(a) Annual Period



(b) High Demand Season



(c) Low Demand Season

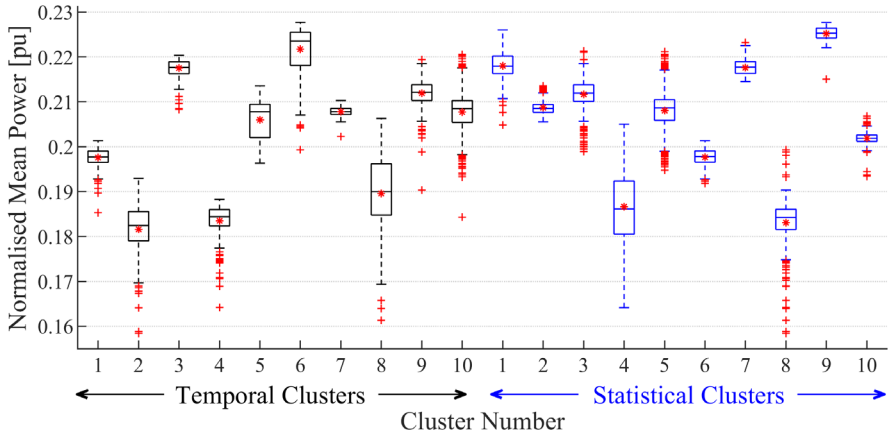
Figure 3.27: The clustered daily solar power profiles depicted for the (a) annual, (b) high and (c) low demand seasons using the ToU feature-based clustering approach.

3.5.4 Statistical comparison and analysis

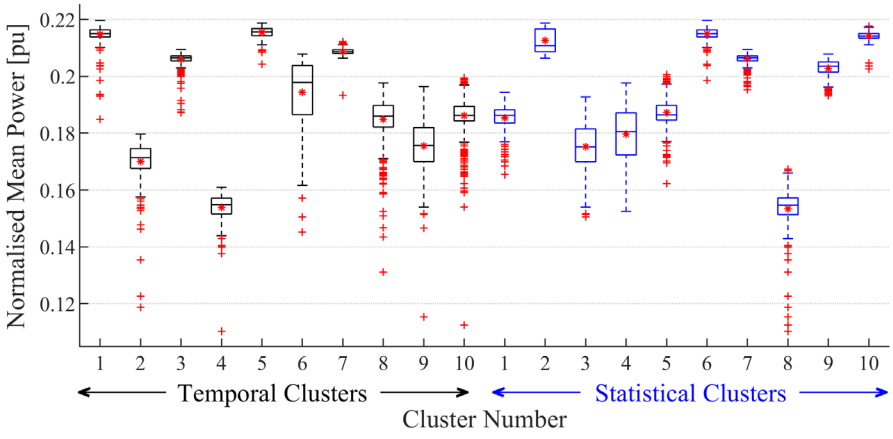
Statistical comparisons of the cluster formations are performed for the temporospatial and the 13 ToU feature-based approaches. This is achieved by comparing the means of the solar power profiles associated with the individual clusters using standard boxplots. In the granularity of the plots, the overall mean solar power of each cluster is indicated by a red dot within each boxplot, the median of each mean cluster

member profiles is indicated by the centre horizontal line, and the bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. The whiskers extend to the extreme data points, excluding the outliers, which are indicated by the tail symbols depicted outside the whisker ranges. The results are presented for the solar power profiles of the full five-year temporal span of the input dataset.

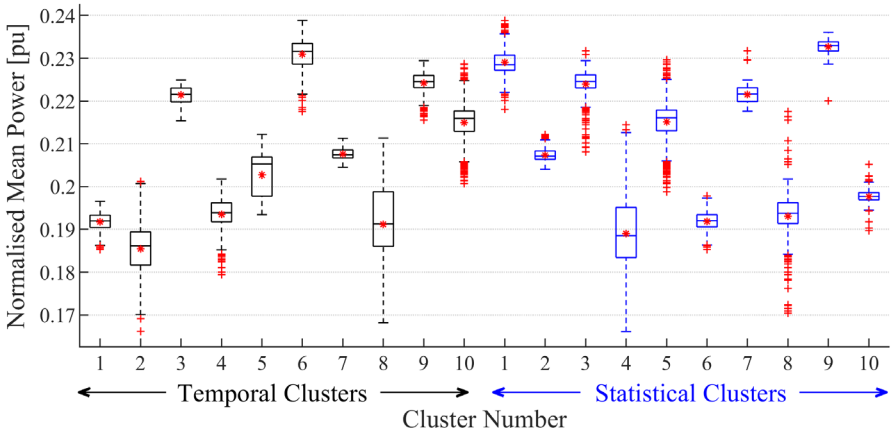
Figure 3.28(a) shows boxplots of the means of the solar power profiles associated with the individual clusters, averaged over the five annual periods. On average, the clusters obtained with the ToU feature vectors depict the cluster with the highest mean value, where the mean values for all clusters have a higher combined mean power. Figure 3.28(b) and Figure 3.28(c) show boxplots of the means of the solar power profiles associated with the individual clusters, averaged over the high demand season and low demand seasons, respectively. In Figure 3.28(b) for ToU clustering approach, the high demand season depicts 5 clusters which show a mean of over 0.2 pu power, while the temporal clusters only formulate four clusters with a mean above 0.2 pu. The same is seen in Figure 3.28(c), where four clusters show mean values above 0.22 pu and only 3 are depicted in using the temporal approach. This shows a bettering in the formulation of the clusters with similar characteristics.



(a) Annual period



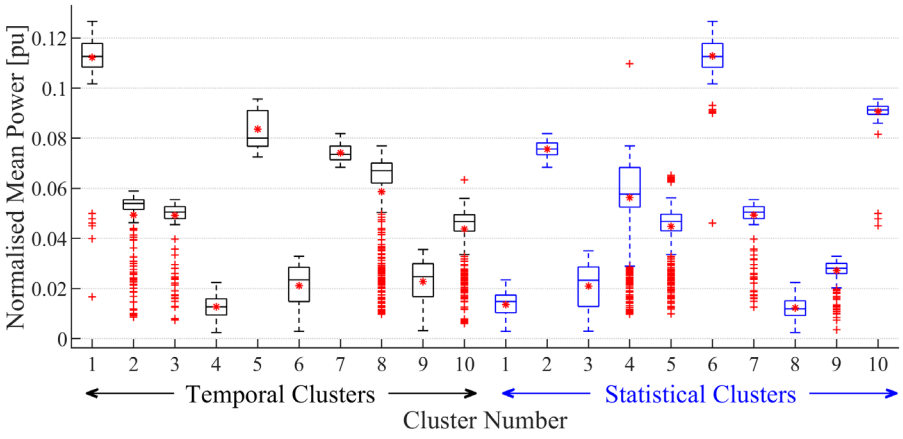
(b) High Demand Season



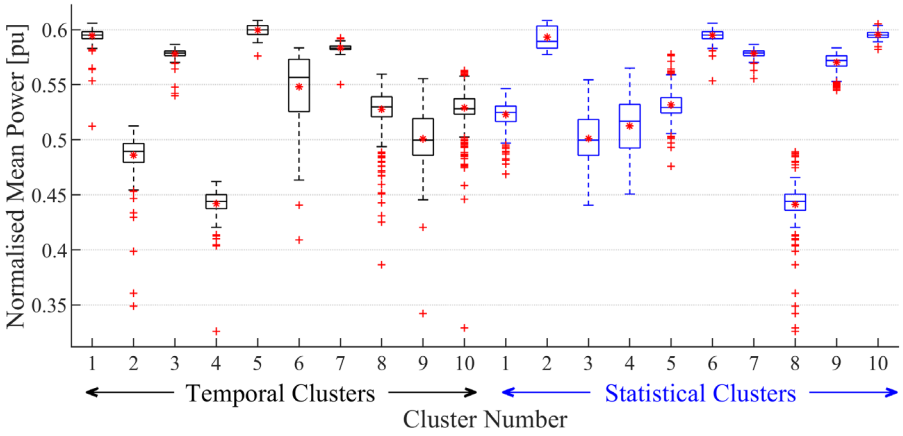
(c) Low Demand Season

Figure 3.28: Boxplots of the normalised mean power of the solar profiles associated with the individual clusters obtained by clustering the temporal profiles and ToU feature vectors of all REDZs, for the (a) annual, (b) high demand season and (c) low demand season averaging intervals.

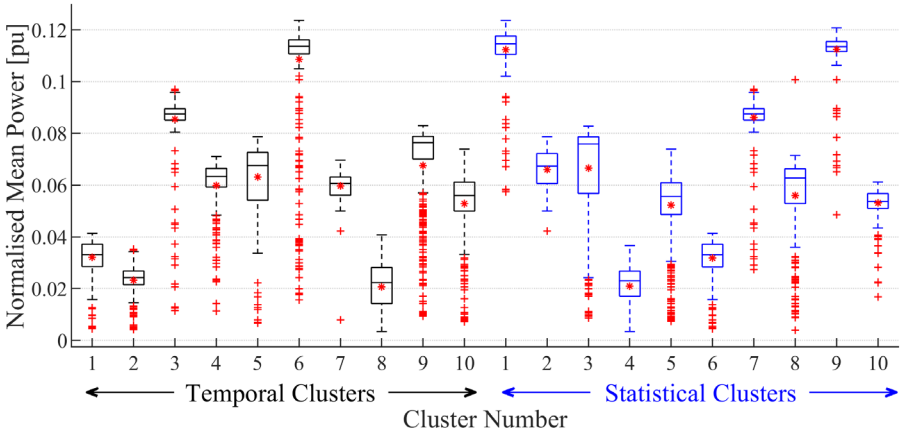
Figure 3.29 and Figure 3.30 depict the (a) morning peak, (b) midday standard and (c) evening peak ToU periods for the high and low demand seasons, respectively. These periods were chosen since the solar bell curve is most dominant during these times periods. In Figure 3.29(c) and Figure 3.30(c), both ToU clustering approaches depict higher mean values, specifically in clusters 1 and 9. When examining Figure 3.24, it is seen that these clusters both belong to the Springbok region, which was divided into two main clusters for the ToU clustering approach and is only represented by one cluster for the temporal approach. Although overall the two clustering methods show similar outputs, the ToU clustering approach shows notably higher means in the peak ToU periods. This is a significant result, since the clustering approach takes cognisance of the periods which are of most value from a System Operators' point of view. The ToU feature based approach also significantly reduces the number of input variables per location, and approach is still able to accurately formulate clusters without degradation of the underlying data characteristics.



(a) Morning Peak Period

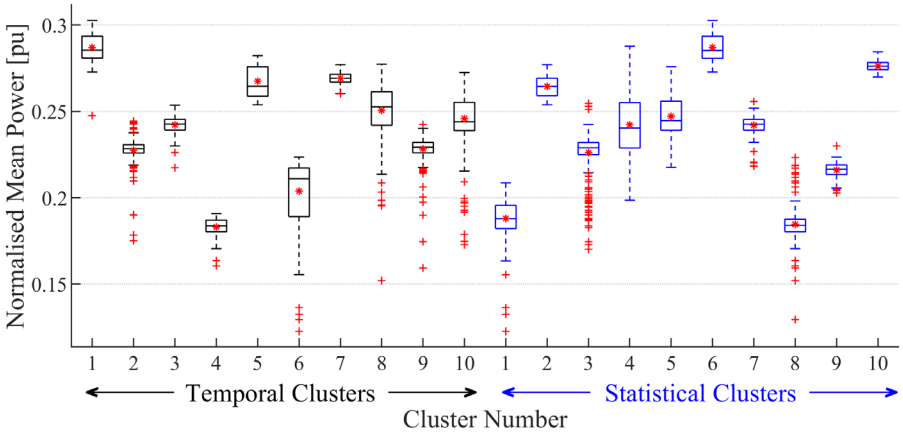


(b) Standard Period

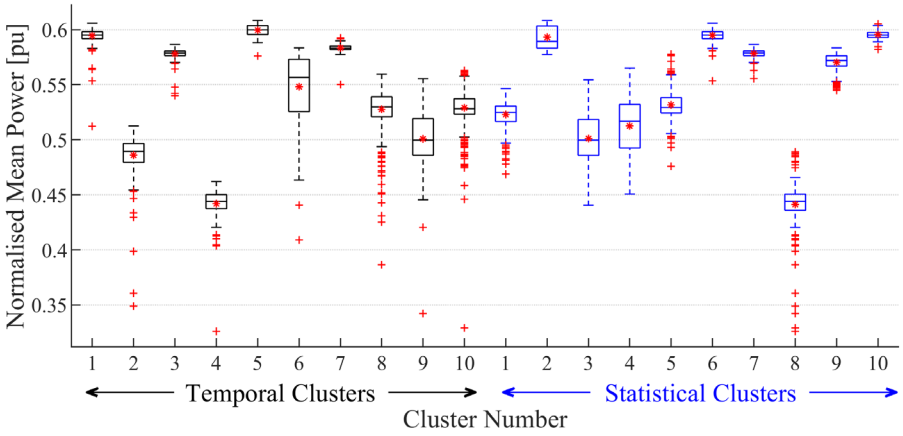


(c) Evening Peak Period

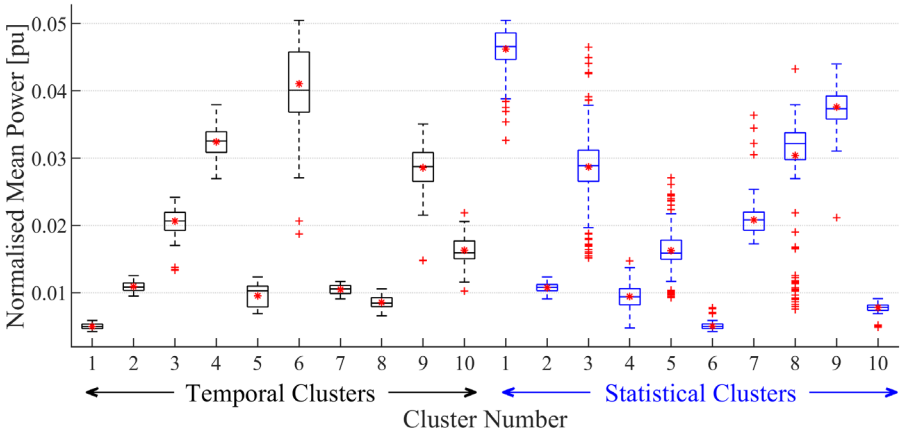
Figure 3.29: High demand season boxplots of the normalised mean power in the (a) peak, (b) standard and (c) off peak periods of the profiles associated with the individual clusters obtained by clustering the temporal profiles and ToU feature vectors of all REDZs.



(a) Morning Peak Period



(b) Standard Period



(c) Evening Peak Period

Figure 3.30: Low demand season Boxplots of the normalised mean power in the (a) peak, (b) standard and (c) off peak periods of the profiles associated with the individual clusters obtained by clustering the temporal profiles and ToU feature vectors of all REDZs.

Table 3.8 compares the ToU feature based approach with the temporal input approach across various metrics. Comparison of the input dataset sizes shows a significant reduction for the ToU feature-based input, with the evaluation metrics yielding superior outcomes for the ToU feature input set for all instances.

Table 3.8: Comparison of the cluster evaluation metrics for the ToU features and the temporal solar power profiles inputs.

Input dataset	Input dataset dimensions	Execution time [seconds]	Calinski-Harabasz	Davies-Bouldin	Silhouette width
ToU features	26 x 4470	5.8645	5129.3971	0.7417	0.6866
Temporal profiles	43 824 x 4470	190.7596	772.2525	1.4686	0.4396

3.5.5 Solar power clustering conclusion

This section explores the translation of spatiotemporal solar power profiles into feature vectors defined in terms of the statistical properties of these profiles for a predefined set of ToU periods. The proposed

approach allows for spatiotemporal solar resource data to be translated to spatial cluster formations, with due cognisance of the temporal characteristics of the demand profile. The resulting clustered resource maps are of value for the siting of solar generation capacity such that the power generation profiles deliver optimised complementarity with the solar resource profiles in the context of the daily and seasonal residual load profile. Furthermore, for the case study of the spatiotemporal solar resource dataset associated with the South African REDZs, the translation to feature vectors reduces the dimensionality of the dataset to be clustered significantly, i.e., from 43 824 hourly solar power observations for each of the 4470 spatial coordinates to 26 ToU feature values per spatial coordinate.

Overall, the results presented in Figure 3.29 and Figure 3.30 show that the clusters can assist in identifying appropriate geographic regions to optimise energy yield and variability targets for specific demand seasons and ToU periods. The clusters obtained with the ToU feature vectors generally rate superior compared to the clusters obtained with temporal profiles. Clustering by ToU periods using the ToU feature vectors effectively partitions locations into clusters with distinctive seasonal solar characteristics. For instance, Cluster 1 obtained with the ToU feature vectors, as shown in Figure 3.30(c), has the highest normalised mean power for all evening peak periods.

The derived cluster maps have application in the optimal spatial allocation of solar generation in the context of grid support. It is evident from the results that clustering based on ToU statistical features presents an opportunity to leverage the temporal diversity of spatiotemporal solar resources

3.6 Conclusion: Resource data clustering

The cluster distributions obtained with the ToU feature vectors are compared with distributions delivered by clustering the spatiotemporal profiles using the k-means algorithm. This delivers a significant result in the sense that the distributions obtained with the ToU feature vectors across the eleven geographical REDZs, for the same number of clusters, show significantly higher granularity and diversity. It is concluded that the clustering of ToU feature vectors, compared to the clustering of raw temporal profiles, distinguishes more finely between the daily and seasonal temporal properties of neighbouring geographic locations. Overall, the results show that the power profiles associated with the individual clusters obtained by the feature vector approach have similar or higher means and similar or lower standard deviations compared to the clusters obtained with spatiotemporal profiles.

The ToU based clustering approach has potential as a data reduction pre-processing step in optimising the locations of future wind and solar energy plants for optimal grid support in the context of the diurnal and seasonal characteristics of the aggregated load profile. The ToU periods used in the case study can be readily amended to represent the operating characteristics of different systems, or incorporate different criteria, such as time-dependent transmission constraints, real-time pricing considerations, etc.

Chapter 4

Optimised spatial capacity allocation framework

Chapter 4 provides a detailed view of the optimised spatial capacity allocation framework, where a comprehensive systematic literature review is completed to determine where the proposed model fits into the existing body of knowledge. The remainder of the chapter details the implementation of the optimisation framework proposed.

4.1 Systematic literature review: Factors influencing the proposed model

This section details factors influencing the formulation of a RE integration strategy to optimise the geospatial capacity allocation of wind and PV resources.

4.1.1 Systematic literature review methodology

The aim of this research is to develop an optimisation framework which addresses the following research question:

> *Can a comprehensive RE geospatial capacity allocation framework be developed that incorporates a risk-based approach to grid support, taking cognisance of real-world regional siting constraints, using geospatial wind and solar resource clusters as the input dataset?*

The expected outcome of this research is to develop a model which determines the number of temporospatial clusters an area should be subdivided into, and then statistically determine the optimal geospatial capacity allocation of the RE fleet to the clustered zones, based on a risk adverse approach.

It is vital to examine, understand and critically engage with the literature that forms the body of knowledge pertaining to the chosen research question. In this study the aim is to conduct a systematic literature review to determine what relevant research publications exist. The main literature search is done through three scientific research databases, namely, the Scopus database[1], Engineering Village database[2] and IEEE Xplore Digital Library[3].

The literature search aims to identify a body of work that shows similarity to and supports the proposed optimised geospatial capacity allocation framework, while simultaneously identifying the research gaps. This is defined in the following reformulated research question:

> *Does a model / framework / roadmap exist that determines the optimised geospatial capacity allocation of clustered wind and solar regions for grid power balance.*

[1] Scopus is an abstract and citation database introduced by Elsevier. It includes publications from various fields namely, social sciences, earth sciences, physical sciences and life sciences.

[2] Engineering Village is a research database with high-quality, cross-disciplinary content, which enables researchers to perform thorough and effective literature reviews as well as analyse the research landscape and solve problems.

[3] IEEE Xplore is a research database for discovery and access to conference proceedings, journal articles, technical standards, and related materials on topics, such as electrical and electronic engineering, computer science, and related fields.

The keywords searched are derived from this research question. The selected keywords are searched for through all paper titles, abstracts and research keywords within the three literature databases.

The search line is as follows:

> *("renewable energy" OR RE) AND (wind AND solar) AND (geographic OR sit*[4] OR location* OR plant* OR farm*) AND (planning OR allocation) AND (optim*) AND (capacity OR mix) AND (peak OR load OR demand OR supply OR balance)*[5]

The Scopus database found 95 record matches, of which 51 are journal articles, 35 are conference articles and 9 are other sources, i.e., conference reviews, reviews and book chapters. The Engineering Village database found 132 record matches, of which 70 are journal articles, 52 are conference articles and 10 are other sources, i.e., conference proceedings, book chapters and articles in press. IEEE Xplore digital library found 73 recorded matches, of which 11 are journal articles and 62 are conference articles. With an emphasis on novelty, the journal articles are made the focus of the literature review, while the conference articles and other literature are briefly scanned for novel contribution. Throughout all databases there is a total of 133 journal articles matches, of which 40 are duplicate studies that were removed. This results in 93 journal records which match the keywords for the formulated research question.

In the initial record screening of the 93 journal articles, 22 documents are deemed irrelevant and excluded. This leaves 71 document abstracts showing varied similarities to the proposed model. The 71 full-text articles are then thoroughly assessed for eligibility and a further 26 full-text articles are excluded with reference to the exclusion criteria. The criteria used for document exclusion include research where the optimisation objectives vary significantly compared to the proposed method. The exclusion criteria also includes isolated and off-grid related research, review journals or reviews of country specific energy practices, research highly dependent on storage optimisation, wind or solar resources seen only as a peripheral constraint, and studies that focus on the transmission network level optimisation of the RE resources. Figure 4.1 depicts a flow diagram of the systematic literature review process including the resource identification, screening, eligibility and inclusion stages.

[4] Inclusion operator, used to find keywords in research which include any variations in the word, which may be filled into the inclusion operator space. E.g., optim* includes optimise/optimise, optimal, optimisation, etc.

[5] Search line including logical operators AND and OR, used to incorporate various key words and phrases into the literature content found. The inclusion of brackets and logical operators allows for the multiple variations of the same concept to be used, as to widen the search space and avoid exclusion of important literature based on varying terminology.

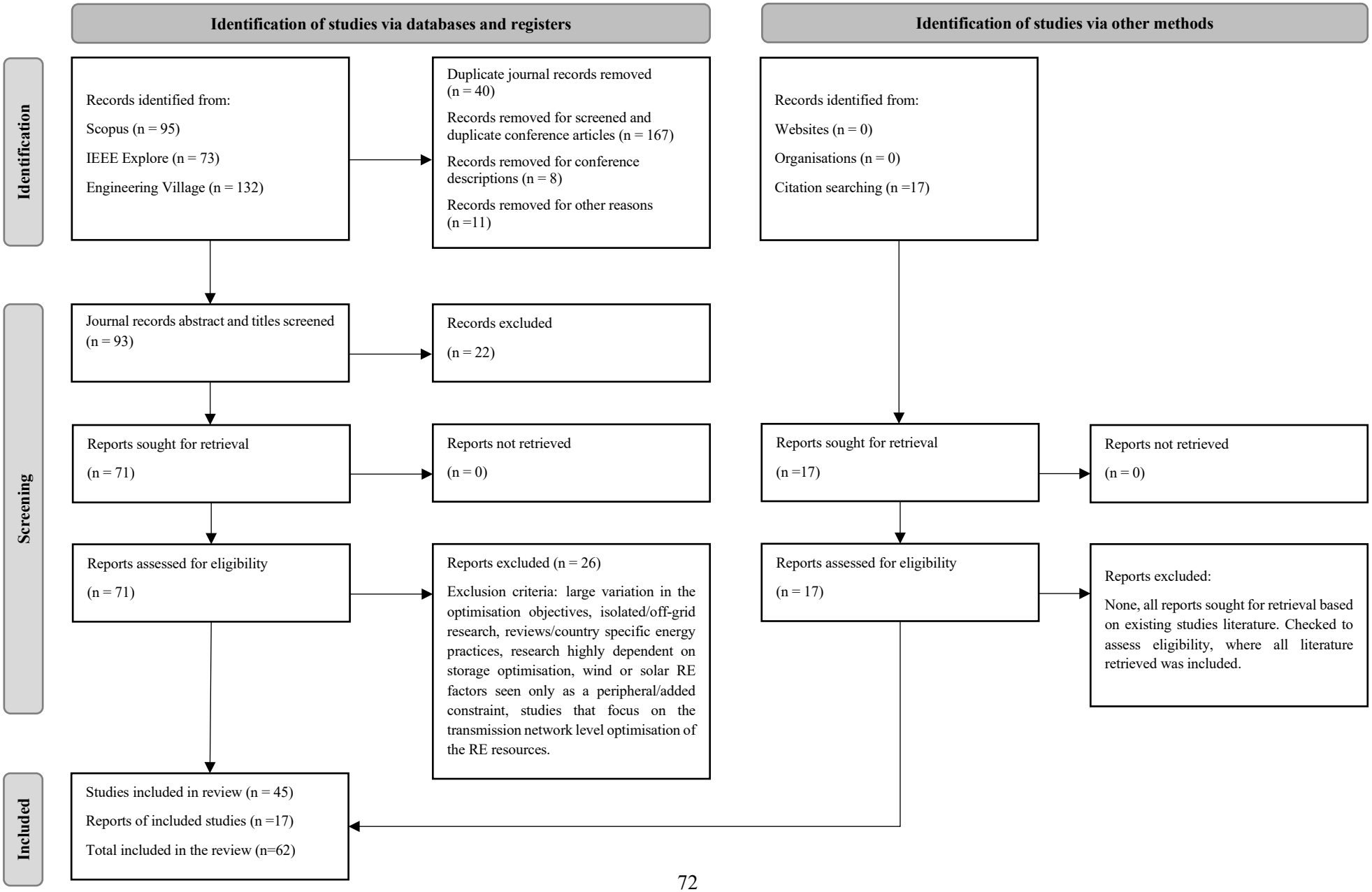


Figure 4.1: Flow Diagram of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) implemented in the systematic literature review.

4.1.2 Review on the existing body of knowledge

4.1.2.1 Overview

Renewable energy sources represent a clean energy alternative to the commonly used fossil fuel energy counterpart. Many countries have set medium to long term planning targets to reach 100 % renewable energy generation systems [120, 121, 122, 123]. However, the integration of this green energy alternative has given rise to increased variability and uncertainty in the power generation mix [124]. Consequently, the resulting residual load profile, i.e., national demand less the renewable generation component, requires increased ramping of conventional generation as well as leveraging of ancillary services to maintain power balance [125].

It is generally accepted that geographic dispersion and correlated geographic siting of renewable energy generation capacity reduces the variability of the aggregated power generation profile [126]. The stochastic nature of the renewable generation sources can thus, to an extent, be managed through optimal geographic capacity allocation and site dispersion of renewable energy generation units.

RE resource integration studies have been conducted worldwide. Studies have been published for Poland [127], India [128], The USA [129, 130, 131, 132], Australia [133, 134] and Europe as a whole [135, 136]. The literature shows that no detailed study has been reported for the planning and integration of the renewable wind and solar energy fleet in South Africa.

Although numerous studies aimed at geospatial optimisation of the siting and sizing of renewable plants have been reported in literature, many of these show limitations, such as the following:

- Complementarity of the RE sources is optimised without inclusion of the demand profile.
- Capacity is allocated geographically for predefined locations rather than optimising the spatial distribution, or the implementation is defined for a limited geographic area.
- Optimise temporally to minimise the RE generation and demand mismatches.
- Minimise correlation coefficient-based objectives.
- Optimisation is based on minimising cost rather than targeting the characteristics of the RE generation profiles.
- Consider energy storage and other ancillary services to predominately meet the demand shortfall / peak demand periods.

The following sections detail the studies assessed based on these defined limitations.

4.1.2.2 Studies investigating complementarity without the demand profile

Several of the studies aimed at exploring the nature of the complementary relationships between spatiotemporal wind and PV power generation profiles focus predominantly on minimising the variance of the resultant RE generation profile [137, 138, 6, 139, 140, 141, 142, 143]. This is done without considering the temporal characteristics of the demand profile. The studies confirm that complementarity between the wind and solar resources can be utilised to create an aggregated generation output that reduces the effect of the stochastic nature of the RE resource. The complementary relationships can be formulated according to the combination of technologies, in this case wind and solar, as well as by considering multiple geospatial locations with varied climatic and temporal characteristics [138, 6, 139, 144, 145].

An early investigation by Takle & Shaw [137] examines the complementary behaviour of the wind and solar resources on an annual and seasonal basis, in Central Iowa USA. The results show high resource

complementarity on an annual basis and only slight complementarity on the daily scale. Widén [138] considers the correlation between large scale integration of wind and solar power, noting the effects of geographic resource dispersion in Sweden. The results show negative correlation between these resources on all time scales, resulting in a significant smoothing effect on the aggregated output of dispersed generation units. Hoicka & Rowlands [6] explore whether the combination of solar and wind resources in selected locations serve to reduce the variability of the power production in Ontario. The study further investigates whether the reduction in variation can be further improved by dispersing the resources geographically, as well as when the number of solar and wind resource locations is increased. The results of this study show that the combination of resources increases the smoothness of the resulting generation profile, and that increased geographic dispersion and an increased number of generating units further reduces the output variability. Santos-Alamillos et al. [139] completed a similar study, where the complementarity of wind and CSP power is assessed. The results of the study confirm that combining geographically diversified resources results in spatiotemporal power balancing.

Other complementarity studies include the optimised ratio of the combined resources [146, 142, 140]. Han et al. [140] considers the complementarity of wind power, photovoltaic power and hydropower, while assessing the optimal ratio of wind and solar resources for power system planning in North China. The complementarity of these resources is investigated by examining the fluctuations of the individual and combined power generation profiles. Ren et al. [146], similarly, assess the complementarity of the wind and solar resources in China, with the aim to reduce the output fluctuations associated with RE generation. The relationship between the wind and solar resources is determined using Kendall correlation, and it is shown that optimising the complementarity of the wind and solar resource fleet decreases the risk of days with zero output power from the variable RE sources, thereby increasing the reliability of the power supply. The return levels of hourly decrease in the wind and solar power output are also significantly decreased by introducing complementary RE sources, which reduces the need for ancillary services. Zhang et al. [142] propose a multi-level framework to assess the synergy of the regional wind and solar resources in Shandong province, China. The framework initially employs clustering to reduce the input data set dimensions while capturing the temporal and spatial attributes of the underlying data set. The cluster sets are used as the input data to capture the complementarity of the wind and solar resources. The mutual complementarity between the wind and solar resources is determined and defined by a local synergy coefficient. A matrix of the anti-correlation patterns of the regional resources is determined and used as the input variable for the clustered region formations. The use of clustered inputs in [142] shows some similarity to the methodology proposed in this study. However, the predominant focus of Zhang et al. is to assess the complementary nature of the wind and solar resource, and demand is not included in the study.

Koivisto et al. [144] investigates the variance of the aggregated renewable power generation profile by optimising the capacity allocations and siting of onshore wind, offshore wind and PV plants for Northern Europe. The incorporation of PV, in conjunction with the geographical dispersion of the RE plants, is shown to contribute to a large reduction in the variability of the RE generation. It is furthermore determined that the optimisation aids in significantly reducing the probability of very high and very low aggregated generation, while reducing large ramping requirements. This is achieved with the minimisation of a single statistical parameter, namely, the standard deviation of the aggregated renewable power. Jerez et al. [145] identifies optimal locations for the siting of wind and solar RE plants, such that the variability of the output generation is decreased while meeting specified efficiency and stability conditions. The implementation of this optimisation model is applied over in the Iberian Peninsula in Spain. In this approach the demand profile is not considered, but rather the RE resource variability is minimised for increasing percentages of the demand profile.

Overall, the above complementarity studies confirm that RE resource diversification and geographic dispersion can be leveraged to decrease the variability of the aggregated wind and solar power generation profile. This represents the basis for optimisation of the RE generation fleet by taking cognisance of the geographic placement and sizing of RE generation units. Several further studies have been conducted to extend this principle by incorporating the demand profile. A large amount of work has been published based on the optimised siting and sizing of renewable energy resources with consideration of the demand profile.

4.1.2.3 Studies investigating complementarity considering the demand profile

Assessing the complementarity of the RE generation profile, including the demand, can be implemented using predefined locations, rather than optimising the RE spatial distributions [123, 147, 148, 129]. This differs from the proposed framework which includes the optimisation of the RE plant placements.

Stoyanov et al. [147] preselected 8 sites to investigate the complementary nature of wind and solar resources with reference to the demand in Bulgaria. The study noted that the optimal RE resource configuration is highly dependent on the energy potential at a given location, as well as the correlation between the wind and solar resources. Jurasz & Mikulik [149] conducted a similar spatial and temporal complementarity study for wind and solar resources in Poland. The study targets seven preselected wind and seven preselected solar sites, where the correlation coefficient is determined for each time step between the RE resource profiles, as well as the demand profile. The objective of the optimisation method employed in the study is to minimise fluctuation in the daily energy yield. The resultant aggregated generation profile is then further investigated with the view to determine the effects of the spatial distribution of the RE sources. The study concluded that increasing the distance between sites increases the correlation coefficient between the temporal RE profiles, which increases the smoothness of the resulting generation output.

Slusarewicz & Cohan [129] studied the complementarity of wind and solar resources for various locations in Texas. Five wind sites and seven solar sites, representing existing RE plants, are considered in the study. The wind and solar resources are studied using two reliability metrics, namely, firm capacity as introduced by Archer & Jacobson [150], and peak average capacity percentage. The capacity factors are computed annually for each site from 2007 to 2013, as well for the summer and winter peak hours. The RE resources are compared over the hourly periods which coincide with peak demand periods, to identify sites that show complementarity with instances of peak demand. The results yield a high percentage of negative correlation when comparing wind and solar resources directly. When comparing the RE capacity with the peak demand periods, the solar resource shows high correlation with the peak demands in summer, and the chosen wind resource locations show high positive correlation with the peak demand in winter. All the above studies consider a limited numbers of RE site locations. In this context, the contribution by Slusarewicz & Cohan acknowledge that there is a need to increase the geographic scope, with the view to identify the exact locations to maximise complementarity and derive the optimised resource capacity mix.

Da Luz & Moura [123] introduce a novel method for determining the optimal complementary mix of wind, hydro, biomass and solar resources to meet the projected hourly and monthly demand profile in Brazil in 2050. The optimisation is specifically aimed at determining a scenario where 100 % of the demand is serviced with RE sources, without experiencing loss-of-load, curtailment of the RE technologies, or requiring the addition of new hydro reservoirs. The study was implemented using spatiotemporal data sets, representative of each technology included in the study. The country is divided into regions based on the annual variations in the characteristics of the wind and solar resources. The study concludes that optimising based on complementarity reduces the energy storage requirements and allows for a higher PV power integration.

There are limited studies that focus on energy storage as a method of meeting the demand shortfalls. However, Hemmati et al. [151] presents a method to smooth the wind and solar resource output as well reduce network losses using energy storage systems.

4.1.2.4 Studies investigating power mismatches between demand and RE generation

A number of studies that focus on the concept of demand matching have been conducted. In these studies, the RE fleet is optimised to reduce the mismatched hours between the resultant RE profile and the demand profile. However, the geographic locations of the potential RE profiles are not clustered as in the methodology proposed in this dissertation. The objective is to reduce the overall variability of the resultant residual load profile, rather than the frequency of occurrence of the maximum residual hourly load values. The studies are, furthermore, implemented using temporal resource profiles, rather than profiles derived using a statistical approach, as is the case for the proposed framework.

Multiple investigations have been conducted to determine the hourly power mismatches between the demand and the RE generation profiles [148, 130, 136], as well as the spatiotemporal reliability of wind and solar resources to meet the total demand [131]. Other studies focus on the maximum penetration levels achievable for wind and solar resources [132], or the optimised ratio of the installed wind and solar resource capacities, by considering their complementarity in the context of demand [135, 152].

Heide et al. [148] developed a weather driven modelling approach to optimise the RE supply based on the hourly power mismatches between the RE generation and load in Europe, for a fully RE generation fleet. The mix of the wind and solar fleet is optimised such that the storage energy capacity, annual balancing energy or balancing power is minimised. Ramirez Camargo et al. [153] explore the time-dependent variability of the RE profile, by presenting a spatiotemporal analysis of potential RE locations within municipal areas in Germany. The study identifies locations that provide an optimised solution for a balanced local energy supply, considering the local demand profile. Clack et al. [130] introduces two novel linear programming approaches to aid in the planning and design of the electric power systems in the USA. The first optimisation approach combines conventional generation types, renewable energy sources and storage methods to minimise the residual load profile. The second optimisation approach aims to minimise the annual system costs, whilst considering the parameters included in the first optimisation. The optimisation method employs a load-matching technique, using a temporal approach, where minimised variance of the residual load is calculated over a desired time interval. The residual load refers to the demand minus the aggregated RE electricity generation. The second approach takes cognisance of the first optimisation approach, with the aim of minimising overall system cost. The load matching technique calculates the shortest distance between the generated electricity and the load for each time interval over a specific temporal span. Both methodologies optimise over the various sites and capacity allocations. The optimisation routines simultaneously include the existing and future expansion of the conventional and variable generators, High Voltage Direct Current transmission, and storage over large geographic area. The results show that both methods provide robust approaches for modelling an optimised electric power system over a large geographic region with high temporal resolution. The cost optimisation approach yields superior results for optimising a real-world power system. Mareda et al. [136] investigate hourly power mismatches between the demand and the RE generation profiles. This approach takes advantage of the geographic resource complementarity between renewable energy sources and energy storage methodologies across Europe. A parameterised optimisation method, using the genetic algorithm, is developed to determine the optimal spatial distribution of the RE units. This optimisation methodology takes into account the renewable energy penetration mix as well as the required reserve power capacity. The study highlights trade-offs pertaining to grid storage, grid extension and the optimal wind and solar mix.

Shaner et al. [131] analyse the spatiotemporal reliability of wind and solar resources to meet the total demand in the US. The temporal characteristics of the wind, solar and demand profiles are examined over a 36-year period. It is concluded that providing a highly reliable energy supply with only wind and solar resources would require a large degree of support technology, including demand management, energy storage, flexible generation and so on. Nikolakakis & Fthenakis [132] investigated the maximum penetration levels achievable for wind and solar resources, combined and individually, as a function of grid flexibility in the state of New York, USA. The optimisation method implemented compares hourly wind and solar resource profiles to demand over an annual period. The proposed penetration level optimisation methodology takes cognisance of flexibility constraints and dumped electricity by setting maximum constraints. The study concludes that combined wind and solar resources yields a higher penetration level with lower curtailment rates. The PV component contributes an additional benefit in the sense that the PV generation profile satisfies the peak demands periods in the summer months, which is associated with higher demand costs.

Zappa and van den Broek [135] explore the optimal integration of wind and solar energy resources in Europe. The study investigates the optimal mix these resources as well as the effects of optimal plant placement for minimised residual demand. The study concludes that the optimised placement of renewables has minimal effect on minimising the residual demand profile within the European context. This work has some synergy with the methodology proposed in this dissertation. However, there are distinct differences in the two approaches, especially in the sense that the study does not employ the initial clustering of the RE resources. The objective function used to reduce the residual load profile also differs, in that the proposed method uses a statistical approach to reduce the expensive high load operating periods. Coutinho et al. [152] determines the optimised ratio of the installed wind and solar resource capacities, by considering their complementarity, for best fit to demand in the Açurua, Brazil. The hourly RE generation profile is fitted to the hourly demand profile, and the mean absolute error is used as the metric to determine the error between each time step. The results provide an optimal wind and PV ratio per region to find the best fit to the energy demand.

A few general optimisation studies incorporate the demand profile with the aim to meet optimisation objectives such as the optimal RE penetration level [154], potential RE capacity integration [155], the optimal capacity mix and minimised curtailment [156], or improving voltage quality and system reliability [157].

Jayapalan et al. [154] implements an optimisation strategy to achieve optimal generation capacity expansion planning until 2044 in India. The optimisation model aims to evaluate the environmental influences, as well as the flexibility of the supply, based on variation in the RE penetration level. Jo et al. [156] investigates the optimal capacity mix of the wind and solar resources installed capacity for a Midwestern state in the USA. The proposed optimisation approach yields a RE optimisation matrix that identifies the hours for which the load can potentially by serviced by an alternative energy source, such as wind and solar. The study, furthermore, focuses strongly on identifying the maximum percentage of wind and solar generation that can be integrated into the total generation capacity without resulting in curtailment of RE generation. Khalid et al. [157] propose an optimisation approach aimed at determining the ideal locations and capacities of non-dispatchable and dispatchable distributed generating units, as well as battery storage systems. The objective function aims to reduce losses and improve voltage quality and system reliability. Quijano et al. [155] introduce the MODERGIS Integrated Simulation's Platform as a RE planning and strategy development tool and presents simulation results for Colombia. This model is all encompassing, with the aim to identify potential RE capacity integration, without the use of installed measurement equipment. The model successfully identifies RE installation potential in selected areas.

4.1.2.5 Studies investigating correlation-coefficient based objective functions

Several studies investigate the complementarity of RE generation profiles, together with demand profile, using correlation coefficients. In these approaches, a figure of merit based on a measure of similarity or dissimilarity between profiles is determined, and the profiles are represented either temporally or statistically. Various correlation metrics are used, including Kendall correlation [158], the Mann-Kendal test spearman coefficient [159], Pearson correlation [160] and the Pearson product-moment correlation coefficient [161]. General correlation is also used to optimise the sizing and placement of RE plants [160] as well as to optimally allocate RE capacity to predefined regions [162]. Canales, et al. [163] introduces a methodology to quantify the partial and the total temporal complementarity between energy sources, where this method can assess complementarity between input vectors on different time scales. The predominate aim is to derive an optimisation model that reduces the variability of the residual load by correlating the aggregated RE generation profile with the demand profile [164, 161, 165]. Although the objectives associated with these studies show some similarity with optimisation framework proposed in this dissertation, the methodologies differ distinctly, especially in terms of the definition of the objective functions.

Horst et al. [158] employs a unit commitment optimisation for increasing levels of variable RE penetration in India. The Kendall correlation factor is used to determine the relationship between the RE sources and the load profile. The results show that an increase in spatial variation aids in smoothing the residual load profile, thereby improving power balance. Jurasz et al. [159] utilise reanalysis datasets from ERA5 to implement a complementarity study between wind and solar resources. The spearman coefficient and the Mann-Kendal test is implemented to assess the complementarity of the RE profiles. It is concluded that the highest correlation occurs on a seasonal scale, while the correlation for daily and hourly observations are virtually negligible. Torres et al. [161] determine the correlation between wind, solar and hydro power generation profiles and demand profile for a case study in Portugal, where the Pearson product-moment correlation coefficient is used. Rosa et al. [160] introduce methodology for optimal sizing and placement of RE resources based on resource complementarity in Brazil. The study implements the Pearson correlation metric to determine the resource complementarity. The dataset is derived for specific automated meteorological measurement station locations, and the solar and hydro resources are initially clustered based on factor analysis using the Principal Component Method. The focus areas of the study include the optimal mix, and the complementarity and the maximum penetration of the RE resources. The optimal mix is initially determined without the influence of the demand profile, and the optimal mix then is re-evaluated with the influence of the demand profile. The objective function implemented in the optimisation study is aimed at reducing the variability of the resultant RE generation profile and minimising the difference between demand and generation.

Li et al. [162] analyse wind and solar resource data for a predefined location in Sydney, Australia. The RE data is correlated against the demand profile with the view to examine the complementary nature between the demand profile and the variable renewable energy generation profile. The conclusion of the study documents strong complementarity between the RE resources, while the level of complementarity between the RE generation profiles and the demand profile depends strongly on the location of the RE plants. Complementarity is defined in terms of the correlation coefficient between two given time series. It is concluded that the combined RE generation profiles can be leveraged to provide better matching of demand. Canales et al. [163] introduces a methodology to quantify the partial and the total temporal complementarity between three energy sources, namely, wind, solar and hydro. This is achieved by combining Euclidean vectors, compromise programming and normalisation correlation techniques. This study provides a robust method for assessing complementarity between input vectors on different time scales.

Guozden et al. [164] investigate an optimisation model that aims to reduce the variability of the residual load by correlating the RE production with the demand profile. The proposed optimisation approach is aimed at the optimal siting of the wind and solar resources, where positive correlation between demand and RE generation is favoured. Solomon et al. [165] investigate the possibility of using negative correlation of wind and solar resource outputs to decrease the storage required to meet the demand in California, USA. The study also determined that the optimal wind and solar mix varies as a function of the RE grid penetration level, where the demand is considered in this study.

4.1.2.6 Studies investigating optimised cost-based objective functions

Another variation in the optimised RE integration planning approach involves minimising the investment and operational costs involved, by optimising the siting of RE sources [166, 167], sizing [168, 169] or both [170, 171, 172, 133]. These studies aim to economically optimise network investments for energy planning and integration over a long-term horizon. These studies typically consider the short-term uncertainties associated with of RE generation, as well as demand growth and resource price uncertainties [133].

He et al. [170] employs a unique method, termed SWITCH-China, to investigate optimised capacity expansion in China. The framework identifies the minimised cost solution for achieving national energy and climate related goals. The optimal capacity mix for 2050 is determined, while the hourly generation dispatch is optimised simultaneously, by taking cognisance of the demand profile. Lenzen et al. [171] simulate a low-carbon electricity supply scenario for Australia with the incorporation of RE generation sources and the demand. The optimisation process determines the locations of RE plants, as well as the expansion of transmission networks. The method represents a cost-based approach, whereby a competitive hourly selection process is simulated to achieve the least-cost configuration. The optimisation output incudes the optimal RE sites as well as the energy mix. Oliveira & Maria [172] proposed an optimisation approach to plan the renewable energy generation for a distribution system in Brazil. The optimisation objective is to determine the optimal placement of the variable renewable resources, with the aim to reduce the investment and operational costs of the RE generation fleet in a long-term planning scenario. The demand profile is incorporated into the proposed optimal planning approach. Simoes et al. [166] investigate the effect of geographical disaggregation of wind and solar resources on the long-term energy system model outcomes for Austria. The study addresses the difference in geographical disaggregation by implementing two scenarios, where low and high disaggregation levels are implemented, and the results are compared. The PV data is divided into five larger regions, with an aggregated profile representing each region. The geographical disaggregation scenarios are analysed based on the system costs incurred to satisfy the demand.

Nunes et al. [133] develop a long-term integration planning approach, with the aim to economically optimise network investments for 2030. A static, stochastic cost minimisation model is employed, by considering the short-term uncertainties of RE resources, as well as load growth and gas price uncertainties expected in the long-term horizon. The study focusses on optimisation of wind, solar and gas resources, by minimising the cost of the system while adhering to policy regulations. The main decision variables are the spatial location, size and RE technology type, as well as requirements for new transmission lines and gas pipelines. Mixed integer linear programming is used as the optimisation solver. Iwamura and Kobayashi [167] proposed an integrated geospatial approach which considers a given wind farm capacity allocation and the associated transmission infrastructure extension costs. This investigation focusses primarily on the optimal siting of wind farms in the context of minimising the total associated capital cost. The constraints imposed on the feasible sites include a minimum capacity factor and the thermal limits of transmission lines. The optimisation strategy uses the Genetic Algorithm (GA).

Candia et al. [173] investigate the effect of large wind and solar energy deployment on the flexibility of the regional power generation system, in Bolivia. The flexibility is measured in terms of energy balancing criteria, by considering the demand profile, power plant scheduling and electricity generation costs. The optimisation model simulates various scenarios with adjusted wind and solar integration levels. The main optimisation objective is to minimise the operational costs, which include, fixed and variable costs, start-up and shut-down costs, ramping and transmission-related costs, as well as loadshedding costs. Pluta et al. [127] explore the impact of increased electricity generation from wind and solar resources over a defined implementation period in Poland. The research questions targeted in the study derives from the Polish energy plan, namely, the *Energy Policy of Poland until 2040*, which aims to increase the RE share to 32 % by 2030. To investigate and validate the feasibility of the implementation of the proposed energy plan, a generalised Unit Commitment and Economic Dispatch approach is used, specifically the Model of Economic Dispatch and Unit commitment for System Analysis (MEDUSA). To verify the long-term plan and the proposed expansion, the suggested framework implements capacity expansion modelling and thereafter employs the MEDUSA. MEDUSA is employed to optimise the operation of controllable power generation units, which is implemented using real technical constraints for an hourly temporal resolution. The objective function employed by the optimisation framework aims to minimise the total cost of the forecasted load balancing, where the dispatchable units are described as in or out of operation at a specific time instance. The costs include operation costs, and the costs of load shifting, demand response and load shedding, as well as start-up and shut down costs. The optimisation includes the centrally dispatched generating units, renewable energy units, the demand, as well as gas units. The result of the study confirms the feasibility of the implementation plan defined by the Polish Energy Policy, stating that the plan is reasonable and consistent with their findings.

Schadler et al. [168] propose a cost optimisation method based on the residual power generation profile for varying wind and solar shares in Germany. The method also identifies the optimal mix of the wind and solar resources, for the scenario that the average renewable energy profile meets 100 % of the average demand profile. The optimisation methodology is implemented using 15-minute temporospatial data sets for the wind and solar resources. The regions for the placement of these RE resources are pre-defined, using the 2-digit ZIP code regions in Germany as the nodes considered. Tróndheim et al. [169] aims to optimise investment and dispatch for energy balance with a 100 % renewable power system by the year 2030 in the Faroe Islands. The study considers storage, as well as the regional demand. The main grid is divided into seven regions, based on the existing transmission grid, together with the smaller Suðuroy region, rather than RE resource characteristics. The optimisation results indicate the exact location and capacity of the added generation, storage, and transmission. The optimisation approach is aimed at minimising the total costs of the electrical power system through a linear optimisation problem based on various capacity and emission constraints.

The studies conducted by Prasad et al. [134] and Priyadharshini et al. [128] focus on methods to reduce the peak demand. Prasad et al. [134] assess the spatiotemporal synergy between wind and solar resources in Australia. The aim of the study is to mitigate the intermittency and variability of the RE resources. Priyadharshini et al. [128] proposes a methodology using wind and solar RE resources to meet the hourly peak demand at minimised cost, for Rameswaram, Tamil Nadu.

4.1.2.7 Studies investigating optimised statistical-based objective functions

Multiple studies have been conducted to evaluate a statistical approach for the optimal allocation and distribution of the RE resource profiles, while considering the demand profile. Certinay et al. [174] proposed a methodology to determine the optimal siting and sizing of wind farm sites in Turkey. The methodology considers both wind resource characteristics, as well as grid constraints and uses a linear optimisation function for maximising the cumulative annual wind power generation, whilst satisfying

the national grid constraints. The proposed approach discards the temporal characteristics of geographic wind speeds by only considering the probabilistic density functions for each location. This methodology only assesses the optimised siting and capacities of wind farms and neglects the inclusion of other RE resource integration.

Vinela & Mortaz [175] determine that by strategically planning for technological and geographical diversification of the RE resource fleet, reduced risk can be achieved for a 100 % RE portfolio in the US. The study considers wind and solar resources, together with demand. Conditional Value-at-Risk optimisation is used as the objective to determine the optimal locations and energy capacities. This is a risk verses return approach, which leverages the variability of the resultant profile against the possible energy return.

Monforti et al. [176] examines the complementarity of wind and solar resources over a single test year in Italy. The demand profile is included, where the hourly, daily and monthly correlation was derived to compute the minimised objective function using a Monte Carlo based approach for the assessment of complementarity over multiple sites. The study is implemented for both short- and long-range spatial complementarity, where both instances showed desirable results for assuring continuous demand coverage. Schindler et al. [177] exploit the complementarity of RE resources to improve the balance between the demand profile and the RE supply in Germany. The results of the study show low complementarity use for wind and solar resources, except on a seasonal scale. Jerez et al. [178] developed a tool that determines an optimised, realistic spatial allocation of given amounts of installed capacities for PV and wind power resources, simulated under varying climate conditions.

There are some similarities between the objectives and methodologies implemented in the above studies and the optimisation framework proposed in this dissertation. However, none of these studies implement the same combination of objective function definition, constraints, optimisation algorithm and clustering approach adopted in the proposed framework.

4.1.2.8 Studies incorporating clustering as an initial data reduction step

A very limited number of optimisation studies have been reported in literature using temporospatial resource clustering as an initial data reduction step.

Siala & Mahfouz [179] applies a novel clustering methodology to initially define optimal regions for energy system planning in Europe. The clusters are determined based on the resource potential of each area. The clustered regions are used as inputs for a modelling framework, termed urbs, where the goal is to minimise the cost of expansion and system operations. The study optimises for both the optimal capacity allocation and the hourly dispatch of generation, transmission, and storage. Couto & Estanqueiro [5] propose a methodology to aid in the expansion of electric power systems by exploiting the complementarity of wind and solar resources to decrease the residual load profile. The method assesses the hybrid power capacity needed to meet the typical daily profiles, the extreme demand values, as well as energy deficits. The method is applied for a Portuguese case study, where the results show that an optimised solar and wind resource fleet increases the stability and sustainability of RE integrated into the electrical grid, as well as reducing the variability and extreme peak values of the residual load profile. The wind and solar resource data are initially clustered based on the resources average power capacity. The current installed capacity is used as a baseline constraint. The objective functions include the minimisation of the residual load as well as the one-hour net load step change standard deviation of the residual load. This method optimises the placement as well as the capacity allocation to different clusters to reduce residual load variability.

Zhang et al. [142] implement principal component analysis and k-means clustering to capture temporal and spatial synergy patterns as well as reduce dataset dimensionality. The synergy patterns are calculated

for all combinations of correlation between wind and solar resources, namely, wind and wind, solar and solar, wind and solar, and solar and wind. This produces four cluster scenarios, dependent on the desired correlation objective. This method differs from the proposed method, as the clustering approach identifies clusters based on the interaction between resource type rather than using the underlying statistical characteristics of the individual resources.

The above optimisation studies exhibit some synergy with the optimisation framework proposed in this dissertation in the sense that they also implement geospatial clustering as a means of categorising the RE sources. However, the clustering approaches used in these studies differ distinctly from the spatiotemporal approaches considered in the proposed framework. Furthermore, there are distinct differences in the optimisation objectives, objective functions and optimisation algorithms implemented.

4.1.3 Overall conclusions on existing literature

Various studies have been conducted to examine the complementarity of the wind and solar resources, without including the demand profile. Where optimisation of the mix and location of these resources are targeted, the objective generally is aimed at reducing the variability of the associated RE power generation profiles, without cognisance of the demand pattern. However, in practice the demand profile is not flat. This implies that the benefits of the optimisation processes do not necessarily translate to an optimised residual load profile. This differs from the framework proposed in this study, where the objective function aims to reduce specific statistical properties of the residual load profile.

The majority of historical studies aimed at optimising the residual load profile generally target a limited number of individual RE locations, which are often pre-selected. The results of these studies yield valuable insights but have limited application in long-term planning where the RE resources span large geographical areas with diverse climatic conditions. In these studies, RE generation capacities are typically assigned to individual locations rather than geospatial regions, as is the case for the proposed framework. This is an important consideration in practice, especially in view of the South African REIPPP scenario, where the locations of RE plants are not only based on plant-specific techno-economic considerations, but also on socio-economic considerations defined on a regional level. Siting for specific coordinates, or a limited number of coordinates, limits the ability of IPPs to select sites that are optimal in terms of practical considerations such as land use, topography, accessibility, cost of land, etc. The proposed optimisation framework addresses these aspects by optimising across diverse spatial clusters.

The studies reported in literature target a wide range of objective functions. These are often defined in terms multiple cost functions including power balance, specific operational aspects such as the use of storage systems, operational costs, the capital cost associated with the RE plants and grid expansion, etc. The framework proposed in this study favours a cascaded optimisation strategy, whereby the residual load profile is optimised statistically to reduce the requirements of ancillary services to complement baseload generation. The operational use of ancillary services, such storage, is thereby considered as a separate subsequent optimisation process.

Several studies investigate the complementarity of RE generation profiles, together with the demand profile, using correlation coefficients. In these approaches, a figure of merit based on a measure of similarity or dissimilarity between profiles is determined and the aim is to combine the RE fleet and the demand profile such that the residual load variability is minimised. This can be done by temporally matching the RE resources to the load or statistically reducing the variability or output ramping, etc. The proposed optimisation framework differs from these strategies in the sense that it adopts a statistical approach aimed at reducing the risk of events associated with high residual load values. The proposed framework, thereby, does not attempt to temporally match the load profile with the RE generation profile.

The literature review identified a limited number of studies where resource clustering has been used as a RE classification tool. However, in these studies the clustering component is generally treated in a cursory manner without considering the suitability of the applied clustering algorithms. Resource clustering represents a core part of the proposed optimisation framework, and the performance of the various clustering algorithms in the context of geospatial optimisation is investigated extensively. From a practical perspective, geospatial clustering is regarded as an essential part of the prosed framework. Apart from the benefit of assigning RE capacities to spatial regions rather than specific coordinates, clustering reduces the dimensions of input dataset dramatically. This reduces the dimensionality of the multi-variable optimisation search space, which translates to reduced risk of local minima and reduced computational cost.

The planning of medium- to long-term RE integration worldwide is mostly conducted using the PLEXOS software platform. This tool typically uses some representative RE resource profile as input. This is a limited approach as the spatial element of the RE resource is not incorporated, and there is a dire need to produce condensed input profiles which represent the diverse characteristics of the underlying RE resource. The proposed integration framework introduces a novel clustering method which drastically reduces the dimensionality of the input dataset, while retaining the underlying properties of the RE resource. The clusters show increased granularity and accuracy when compared to general clustering methodologies.

The research reported in literature, with a few exceptions that is applicable for specific case study scenarios, confirm that geospatial optimisation of RE capacity allocation, especially in the context of the spatiotemporal properties of the RE resources, can contribute to support power balance and various other such operational objectives at high levels of RE penetration. Overall, it is concluded that, compared to the body of work reported in literature, the proposed RE geospatial capacity allocation framework proposed in this dissertation exhibits aspects of novelty and originality, especially in terms of the pre-classification of the RE resource data, the risk-based objective function and the use of adaptable optimisation constraints to accommodate real-world scenarios.

4.2 Spatial capacity allocation framework methodology

4.2.1 Overview

Figure 4.2 shows a high-level block diagram of the proposed geospatial capacity allocation optimisation framework. The processes comprising the framework can be summarised as follows:

- The spatiotemporal mesoscale wind speed and GHI profiles, given by $\boldsymbol{W}$ and $\boldsymbol{S}$, respectively, are converted to power profiles and normalised. This yields the normalised wind and solar power profiles $\boldsymbol{P}^{w}$ and $\boldsymbol{P}^{s}$, respectively.
- The ToU feature vectors of the normalised temporal power profiles are derived and clustered.
- The sets of temporal power profiles representing the wind and solar clusters, given by $\boldsymbol{P}^{cw}$ and $\boldsymbol{P}^{cs}$, respectively, are derived.
- The sets of mean temporal power profiles of the wind and solar clusters are derived, given by $\overline{\boldsymbol{P}}^{cw}$ and $\overline{\boldsymbol{P}}^{cs}$, respectively.
- A statistical objective function F^{o}, defined in terms of the histogram of the residual load profile, is calculated.

– The optimisation algorithm allocates the specified wind and solar generation capacities, given by P_c^w and P_c^s, respectively, to the clusters such that the cost function F^o is minimised. This yields the weight vectors, given by $\boldsymbol{X}_c^w$ and $\boldsymbol{X}_c^s$, whereby P_c^w and P_c^s are allocated to the individual wind and solar clusters, respectively.

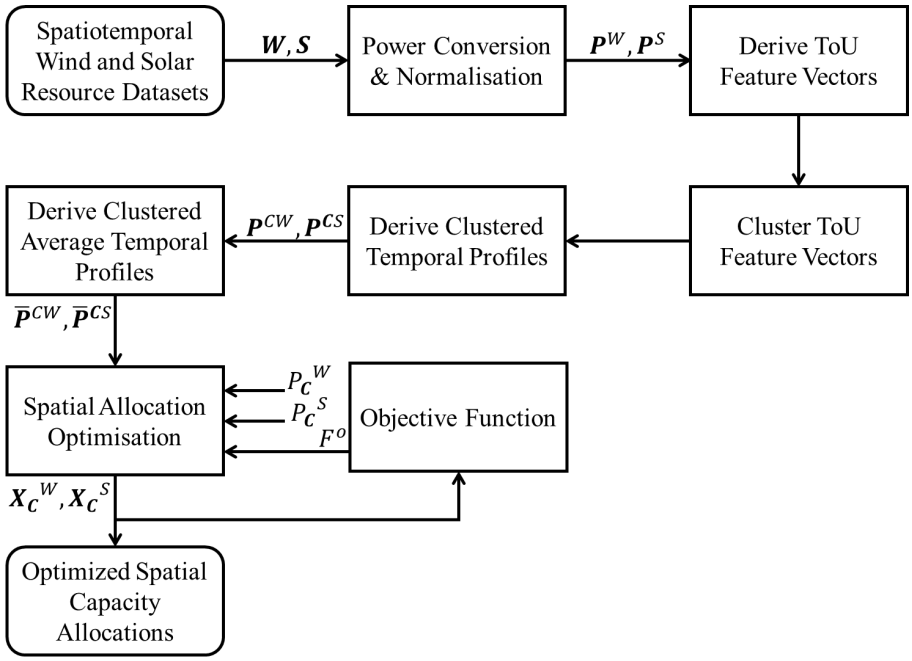


Figure 4.2: Geospatial capacity allocation framework.

4.2.2 Resource classification

The initial step in the optimisation framework involves clustering the RE resources. Clustering translates the power profiles associated with $\boldsymbol{P}^w$ and $\boldsymbol{P}^s$ into clusters with similar temporal properties, each representing a subregion of the geographical target space. Clustering also represents a dataset reduction methodology, which considerably reduces the computational expense of spatial optimisation. In practice, clustered resource maps are also more appropriate for plant siting studies compared to resource maps defined in terms of single coordinates. This is due to the many considerations and constraints that apply for plant siting in practice. With allocation to single sites, there is a high risk of choosing coordinates that are not viable for plant placement. It also limits the ability of IPPs to find the most appropriate sites from plant-specific perspectives such as capital expenditure, terrain topography, accessibility and transport infrastructure, etc.

The power profiles associated with $\boldsymbol{P}^w$ and $\boldsymbol{P}^s$ are clustered using the k-means algorithm. Two clustering approaches are considered, namely, clustering the temporal profiles, and clustering based on the statistical properties of the temporal profiles for ToU periods. The latter approach has been shown to yield superior results for the clustering of spatiotemporal power compared to the clustering of raw temporal profiles [180], and is discussed in detail in Chapter 3.

The resultant wind and solar clusters can be defined mathematically, where the set of wind power clusters $\boldsymbol{P}^{cw}$ is defined by the relationships

$$\boldsymbol{P}^{cw} = \{\boldsymbol{P}_m^{cw} \mid m = 1,2,3 \cdots N_m^w\}, \tag{4.1}$$

$$\boldsymbol{P}_m^{cw} = \{\boldsymbol{P}_{mn}^{cw} \mid n = 1,2,3 \cdots N_{mn}^w\} \tag{4.2}$$

and

$$\boldsymbol{P}_{mn}^{cw} = \{p_{mn}^{cw}(t_k) \mid k = 1,2,3 \cdots N_k\}, \quad 4.3$$

where $\boldsymbol{P}_{m}^{cw}$ denotes the set of power profiles associated with the m^{th} cluster, N_m^w denotes the number of wind profile clusters, $\boldsymbol{P}_{mn}^{cw}$ denotes the n^{th} profile associated with the m^{th} cluster, N_{mn}^w denotes the number of profiles associated with the m^{th} wind power cluster, and $p_{mn}^{cw}(t_k)$ denotes the wind power for the k^{th} sampling interval. The set of mean temporal power profiles $\overline{\boldsymbol{P}}^{cw}$ for the clusters are subsequently derived, given by

$$\overline{\boldsymbol{P}}^{cw} = \{\overline{\boldsymbol{P}}_{m}^{cw} \mid m = 1,2,3 \cdots N_m^w\}, \quad 4.4$$

$$\overline{\boldsymbol{P}}_{m}^{cw} = \{\bar{p}_{m}^{cw}(t_k) \mid k = 1,2,3 \cdots N_k\} \quad 4.5$$

and

$$\bar{p}_{\boldsymbol{m}}^{\boldsymbol{cw}}(t_k) = \frac{1}{\mathrm{N}_{mn}^{w}} \sum_{\mathrm{n}=1}^{\mathrm{N}_{mn}^{w}} p_{mn}^{cw}(t_k), \quad 4.6$$

where $\overline{\boldsymbol{P}}_{m}^{cw}$ denotes the mean wind power profile of the m^{th} cluster.

The set of solar power clusters $\boldsymbol{P}^{cs}$ is defined similarly by the relationships

$$\boldsymbol{P}^{cs} = \{\boldsymbol{P}_{m}^{cs} \mid m = 1,2,3 \cdots N_m^s\}, \quad 4.7$$

$$\boldsymbol{P}_{m}^{cs}\{\boldsymbol{P}_{mn}^{cs} \mid n = 1,2,3 \cdots N_{mn}^s\} \quad 4.8$$

and

$$\boldsymbol{P}_{mn}^{cs} = \{p_{mn}^{cs}(t_k) \mid k = 1,2,3 \cdots N_k\}, \quad 4.9$$

where $\boldsymbol{P}_{m}^{cs}$ denotes the set of power profiles associated with the m^{th} cluster, N_m^s denotes the number of solar profile clusters, $\boldsymbol{P}_{mn}^{cs}$ denotes the n^{th} profile associated with the m^{th} cluster, N_{mn}^s denotes the number of profiles associated with the m^{th} solar power cluster, and $p_{mn}^{cs}(t_k)$ denotes the solar power for the k^{th} sampling interval. The set of mean temporal power profiles $\overline{\boldsymbol{P}}^{cs}$ for the clusters are subsequently derived, given by

$$\overline{\boldsymbol{P}}^{cs} = \{\overline{\boldsymbol{P}}_{m}^{cs} \mid m = 1,2,3 \cdots N_m^s\}, \quad 4.10$$

$$\overline{\boldsymbol{P}}_{m}^{cs} = \{\bar{p}_{m}^{cs}(t_k) \mid k = 1,2,3 \cdots N_k\} \quad 4.11$$

and

$$\bar{p}_{\boldsymbol{m}}^{\boldsymbol{cs}}(t_k) = \frac{1}{N_{mn}^{s}} \sum_{n=1}^{N_{mn}^{s}} p_{mn}^{cs}(t_k), \quad 4.12$$

where $\overline{\boldsymbol{P}}_{m}^{cs}$ denotes the mean solar power profile of the m^{th} cluster.

4.2.3 Geospatial capacity allocation

The objective of the optimal geospatial capacity allocation strategy is to divide a given wind generation capacity P_c^w and solar generation capacity P_c^s amongst the wind and solar clusters, respectively, such that a predefined objective function is optimised. These wind and solar capacity allocations are represented by the weight vectors $\boldsymbol{X}^w$ and $\boldsymbol{X}^s$, respectively. The weight vector $\boldsymbol{X}^w$ for the wind capacity allocations is defined by the relationships

$$\boldsymbol{X}^{w} = \{x_m^w \mid m = 1,2,3 \cdots N_m^w\}, \tag{4.13}$$

$$0 \leq x_m^w \leq 1 \tag{4.14}$$

and

$$\sum_{m=1}^{N_m^w} x_m^w = 1. \tag{4.15}$$

The weight vector $\boldsymbol{X}^{s}$ for the solar capacity allocations is defined similarly by the relationships

$$\boldsymbol{X}^{s} = \{x_m^s \mid m = 1,2,3 \cdots N_m^s\}, \tag{4.16}$$

$$0 \leq x_m^s \leq 1 \tag{4.17}$$

and

$$\sum_{m=1}^{N_m^s} x_m^s = 1. \tag{4.18}$$

The normalised aggregated wind generation profile $\boldsymbol{P}^{wa}$, which has been optimised, is given by the relationships

$$\boldsymbol{P}^{wa} = \{p^{wa}(t_k) \mid k = 1,2,3 \cdots N_k\} \tag{4.19}$$

and

$$p^{wa}(t_k) = \sum_{m=1}^{N_m^w} \frac{P_c^w}{P_{max}^d} x_m^w \, \bar{p}_m^{cw}(t_k), \tag{4.20}$$

where P_c^w is given in GW.

Similarly, the normalised aggregated solar generation profile $\boldsymbol{P}^{sa}$ is given by the relationship

$$\boldsymbol{P}^{sa} = \{p^{sa}(t_k) \mid k = 1,2,3 \cdots N_k\}, \tag{4.21}$$

and

$$p^{sa}(t_k) = \sum_{m=1}^{N_m^s} \frac{P_c^s}{P_{max}^d} x_m^s \, \bar{p}_m^{cs}(t_k). \tag{4.22}$$

where P_c^s is given in GW.

The wind and solar spatial capacity allocations, given by the weight vectors $\boldsymbol{X}^{w}$ and $\boldsymbol{X}^{s}$ are obtained by implementing a constrained hybrid optimisation strategy using the genetic algorithm and gradient-based nonlinear multivariable algorithm to minimise an objective function F^{O}.

The objective functions considered for performance evaluation of the proposed geospatial capacity allocation framework are defined in terms of the normalised residual load profile. The residual load profile $\boldsymbol{P}^{r}$ is defined by the relationships

$$\boldsymbol{P}^{\boldsymbol{r}} = \{p^r(t_k) \mid k = 1,2,3 \cdots N_k\} \tag{4.23}$$

and

$$p^{\boldsymbol{r}}(t_k) = p^d(t_k) - p^{wa}(t_k) - p^{sa}(t_k), \tag{4.24}$$

where $p^r(t_k)$, $p^d(t_k)$, $p^{wa}(t_k)$ and $p^{sa}(t_k)$ denote the normalised residual load, demand, aggregated wind power and aggregated solar power at time t_k, respectively.

The optimisation objective functions proposed in this study can be summarised as follows:

Objective function 1: The objective function F^O is defined as the number of residual load values above a given threshold P_T. This is expressed as

$$F_1^O = \sum_{k=1}^{N_k} \begin{cases} 1, & p^r(t_k) \geq P_T \\ 0, & p^r(t_k) < P_T \end{cases}. \quad 4.25$$

This objective function targets the probability density function of the residual load profile in the sense that it seeks to minimise the number of events featuring higher hourly residual load values.

Objective function 2: The objective function is defined as the cumulative sum of the residual load values above the threshold P_T. This is expressed as

$$F_2^O = \sum_{k=1}^{N_k} \begin{cases} p^r(t_k), & p^r(t_k) \geq P_T \\ 0, & p^r(t_k) < P_T \end{cases}. \quad 4.26$$

This function is similar to objective function 1 but represents a linear weighted approach whereby higher load values contribute comparatively more to the objective function compared to lower load values. The function can be readily adapted to implement non-linear weightings, such as an exponentially increasing weighting function, which would be more effective in targeting extreme values.

The objective functions employed target the higher load values in the residual load profile histogram in a statistical manner. As such, it represents a probabilistic risk-based approach that seeks to minimise the number of events where high residual load values require ancillary service interventions to maintain power balance. From a financial perspective, the proposed approach also reduces the capital expenditure requirements associated with baseload generation capacity and ancillary service, as well as the energy costs associated with expensive peaking generation, such as open-cycle gas turbines.

In practice, the objective function can be easily amended to target the probability density function of the residual load profile for a given ToU period, etc.

In practice, optimal siting of RE generation is not only defined in terms of power balance, but is subject to considerations such as availability of grid infrastructure, power evacuation capacity constraints, economic and socio-economic considerations, land use and availability, environmental impacts, etc. [176, 133]. Together, these considerations translate to minimum and maximum RE generation capacity constraints for a given cluster or geographical area. These constraints should be taken into consideration in a robust spatial capacity allocation framework. These capacity allocation constraints can be formulated mathematically in terms of the weight vectors. The weight vector $\boldsymbol{X}_c^w$ for constrained wind capacity allocations at cluster level can be represented by the relationships

$$\boldsymbol{X}_c^w = \{x_m^w \mid m = 1,2,3 \cdots N_m^w\}, \quad 4.27$$

$$P_{m\,min}^w \leq P_c^w x_m^w \leq P_{m\,max}^w \quad 4.28$$

and

$$\sum_{m=1}^{N_m^w} x_m^w = 1, \quad 4.29$$

where $P_{m\,min}^w$ and $P_{m\,max}^w$ denote the minimum and maximum wind capacities to be assigned to the geographical area associated with cluster m, respectively. The weight vector $\boldsymbol{X}_c^s$ for constrained solar capacity allocations at cluster level can be represented similarly by the relationships

$$\boldsymbol{X}_c^s = \{x_m^s \mid m = 1,2,3 \cdots N_m^s\}, \quad 4.30$$

$$P_{m\,min}^s \leq P_c^s x_m^s \leq P_{m\,max}^s \quad 4.31$$

and

$$\sum_{m=1}^{N_m^s} x_m^s = 1, \quad 4.32$$

where $P_{m\,min}^s$ and $P_{m\,max}^s$ denote the minimum and maximum solar capacities to be assigned to the geographical area associated with cluster m, respectively.

The minimum constraints can be used effectively to accommodate capacity allocations that are predefined based on criteria such as regional development targets and socio-economic considerations, as is currently applicable for the REIPPP programme. Furthermore, in practice, existing generation capacities in the geographical clusters also translates to minimum constraints in the weight vectors. In these implementations the capacities of the existing plants are then added to the new capacities to formulate capacities to be allocated, i.e., P_c^w and P_c^s.

The maximum capacity constraints can be used to accommodate criteria such as geographic grid capacity constraints, as outlined in the recent Generation Connection Capacity Assessment (GCCA) study [181]. These constraints can also be used to limit the allocations to the regions based on technical criteria such as the distance to load centres, and socio-economic considerations aimed at just transition to RE, whereby excessive benefit by some communities can be avoided.

In practice, both the minimum and maximum constraints are typically defined for geographical regions, such as the REDZs in this study, rather than for the individual geospatial clusters. In order to implement a rigorous approach to handle such constraints, the constraints must be translated to apply to subsets of the geospatial clusters. This can be done by splitting the clusters into subclusters, where each subcluster is associated with a predefined geographical region. The capacities are assigned to the individual subclusters.

Mathematically, the weight vector $\boldsymbol{X}_c^w$ for constrained wind capacity allocations can be redefined such that

$$\boldsymbol{X}_c^w = \{\boldsymbol{X}_m^w \mid m = 1,2,3 \cdots N_m^w\}, \quad 4.33$$

$$\boldsymbol{X}_m^w = \{x_{mr}^w \mid m = 1,2,3 \cdots N_m^w,\ r = 1,2,3 \cdots N^r\}, \quad 4.34$$

$$P_{r\,min}^w \leq \sum_{m=1}^{N_m^w} P_c^w x_{mr}^w \leq P_{r\,max}^w \quad 4.35$$

and

$$\sum_{r=1}^{N^r} \sum_{m=1}^{N_m^w} x_{mr}^w = 1, \quad 4.36$$

where $P_{r\,min}^w$ and $P_{r\,max}^w$ denote the minimum and maximum wind capacities to be assigned to the r^{th} geographical region, N^r denotes the number of geographical regions, P_c^w denotes the wind capacity to be allocated, and x_{mr}^w denotes the weight assigned for the m^{th} wind cluster in the r^{th} region.

The weight vector $\boldsymbol{X}_c^s$ for constrained solar capacity allocations is defined similarly, such that

$$\boldsymbol{X}_c^s = \{\boldsymbol{X}_m^s \mid m = 1,2,3 \cdots N_m^s\}, \quad 4.37$$

$$\boldsymbol{X}_m^s = \{x_{mr}^s \mid m = 1,2,3 \cdots N_m^s,\ r = 1,2,3 \cdots N^r\}, \quad 4.38$$

$$P^s_{r\,min} \leq \sum_{m=1}^{N^s_m} P^s_c x^s_{mr} \leq P^s_{r\,max} \quad 4.39$$

and

$$\sum_{r=1}^{N^r} \sum_{m=1}^{N^s_m} x^s_{mr} = 1, \quad 4.40$$

where $P^s_{r\,min}$ and $P^s_{r\,max}$ denote the minimum and maximum solar capacities to be assigned to the r^{th} geographical region, P^s_c denotes the solar capacity to be allocated, and x^s_{mr} denotes the weight assigned for the m^{th} solar cluster in the r^{th} region.

Some regional constraints, such as the maximum power that can be evacuated from region due to grid constraints, are specified in terms of the total wind and solar power that can be allocated to that region. In these cases, Equations (4.35) and (4.39) are combined to yield

$$P_{r\,min} \leq \sum_{m=1}^{N^w_m} P^w_c x^w_{mr} + \sum_{m=1}^{N^{ws}_m} P^s_c x^s_{mr} \leq P_{r\,max}, \quad 4.41$$

where $P_{r\,min}$ and $P_{r\,max}$ denote the minimum and maximum capacities to be assigned to the r^{th} geographical region.

If any of the constraints are specified at regional level, the above mathematical formulations imply that all clusters must be subclustered in terms of regions. Clusters and subclusters located outside of the regions for which constraints are defined are therefore grouped together in an additional region for which no regional constraints apply. For a subcluster that is not geospatially present in region m, x^w_{mr} or x^s_{mr} is pre-assigned a weigh of zero and is omitted from the weight vector presented to the optimisation algorithm.

In practice, the constraints may translate to a combinational permutation of the above formulations, consisting of minimum and/or maximum constraints formulated individually for the wind and solar allocations at cluster level, minimum and/or maximum constraints formulated individually for the wind and solar allocations at regional level, and minimum and/or maximum constraints formulated for the combined wind and solar allocations at cluster level and/or regional level.

The above approach can handle a complex set of capacity allocation constraints in a robust manner. However, constraints that are defined at a regional level require that the optimisation algorithm take cognisance of the individual subclusters present in those regions. This increases the dimensions of the search space, which gives rise to longer simulation times and increased challenges with problems related to phenomena such as local minima.

4.2.4 Optimisation algorithm

Multiple methods have been implemented for the optimisation of RE resource integration. Some such methods include Conditional Value-at-Risk [182], linear programming techniques [130, 165, 133, 173], nonlinear programming techniques e.g. sequential quadratic programming [123, 160], as well as nonlinear mixed integer optimisation stochastic programming [151]. Other optimisation approaches include the greedy algorithm [5] and the simulated annealing algorithm [145].

Further common optimisation methodologies implemented include various evolutionary algorithms, which have been developed for the optimisation of complex problems [183, 184].

The evolutionary algorithms include the bio-inspired metaheuristic, called artificial immune system [172], particle swam optimisation [151, 185] and the genetic algorithm [186, 187, 188, 136]. The genetic

algorithm is a well-documented and widely used method for the optimisation of RE integration models [183, 189].

The optimised geospatial capacity allocation framework is therefore implemented using a hybrid GA-fmincon approach. The genetic algorithm is a meta-heuristic, population-based algorithm inspired by the theory of natural selection [190, 191]. A population of potential solutions, or individuals, evolve with each generation as new solutions are *born* into the population space, while others *die out* of the population space. The potential solutions improve with each generation as the general fitness or accuracy of the population improves. A selection process identifies the best individuals for *breeding* with other individuals based on their fitness values. This is done to produce new candidate solutions within the population space. Random mutation is also incorporated into the algorithm to introduce variation in the parameters of individuals. The algorithm is repeated until the desired termination criterion is achieved. Once the algorithm terminates, the candidate solution with the highest fitness value in the population is selected as the solution to the problem. In the hybrid approach, the best candidate solution becomes the input into the nonlinear programming algorithm, fmincon, to maximise the likelihood of reaching the global minimum.

4.2.5 Performance metrics

The metrics used in literature to quantify the performance of optimised spatial capacity allocation of RE generation vary widely and are highly dependent on the case study objectives. In this study, three metrics are derived to evaluate the performance of the spatial capacity allocation results. These include metrics which calculate the normalised weighted and unweighted cumulative frequency of occurrence of residual power values above a selected demand threshold, as well as the frequency of occurrence within 10 % of the maximum residual load value.

The normalised unweighted cumulative frequency of occurrence metric, F_1^T, represents the cumulative frequency of occurrence of residual power values above a set threshold P_T, as a percentage of the total number occurrences, N_k. The metric is defined mathematically as

$$F_1^T = \frac{100}{N_k}\sum_{k=1}^{N_k}\begin{cases} 1, & p^r(t_k) \geq P_T \\ 0, & p^r(t_k) < P_T \end{cases}, \tag{4.42}$$

where $p^r(t_k)$ denotes the mean residual power for the t_k^{th} interval.

The normalised weighted cumulative frequency of occurrence metric, F_2^T, represents the cumulative energy above a set threshold P_T, as a percentage of the total energy. The metric is defined mathematically as

$$F_2^T = \frac{100}{\sum_{k=1}^{N_k} p^r(t_k)}\sum_{k=1}^{N_k}\begin{cases} p^r(t_k), & p^r(t_k) \geq P_T \\ 0, & p^r(t_k) < P_T \end{cases}. \tag{4.43}$$

The final metric, F_3^T, represents the frequency of occurrences within a given per unit range P_F^T of the maximum residual load value. This metric quantifies the most expensive operating time when replaced by ancillary services. The metric is defined mathematically as

$$F_3^T = \sum_{k=1}^{N_k}\begin{cases} 1, & p^r(t_k) \geq P_F^T \cdot p_{max}^r \\ 0, & p^r(t_k) < P_F^T \cdot p_{max}^r \end{cases}, \tag{4.44}$$

where p_{max}^r denotes the maximum residual load value. For purposes of the case studies conducted in this investigation, the P_F^T is chosen arbitrarily as 0.90.

4.3 Implementation and case study results

4.3.1 Overview

The optimised spatial capacity allocation framework is implemented for a range of case studies, which are chosen to emphasise the benefits and drawbacks of the proposed methodology. Table 4.1 summarises the case studies explored for the results presented in this dissertation.

Table 4.1: Summary of case studies.

Case study	Objectives
1	Unconstrained capacity allocation scenario
2	Minimum constrained capacity allocation scenario: existing wind and solar capacity
3	Minimum and maximum constrained capacity allocation scenario: existing wind and solar capacity and generation connection capacity constraints [181].

The projected demand for the medium term is uncertain, and the Integrated Resource Plan for 2019 (IRP2019) [14] expects that demand growth is likely to be lower than forecasted. For the purposes of the case studies conducted in this investigation, the wind and solar generation capacity figures projected in IRP2019 are scaled to the demand profile shown in Figure 2.7. The projected wind and solar integration percentage is kept constant, but the total demand capacity is defined as the current load profiles maximum demand value, 36.371 GW. The IRP2019 plans for the integration of 10.52 % solar and 22.53 % wind percentage of the total installed capacity. This translates to approximately 8.2 GW of wind generation capacity for P_c^w and 3.8 GW of solar generation capacity for P_c^s, respectively.

Figure 4.3 depicts a histogram of the normalised national demand profile of South Africa for 2010 to 2014. The figure shows the frequency of occurrence of demand power values during the peak, standard and off-peak periods over a range of normalised demand power. The histogram is normalised to the number of hourly time instances within the five-year period. The maximum frequency of occurrence occurs at approximately 0.82 pu power and the second highest frequency of occurrence at approximately 0.64 pu power. The peak and standard ToU periods occur predominantly in the higher power range above 0.8 pu, while the off-peak period is dominant below 0.77 pu power. The optimisation objective function targets the frequency of occurrence of the higher load values

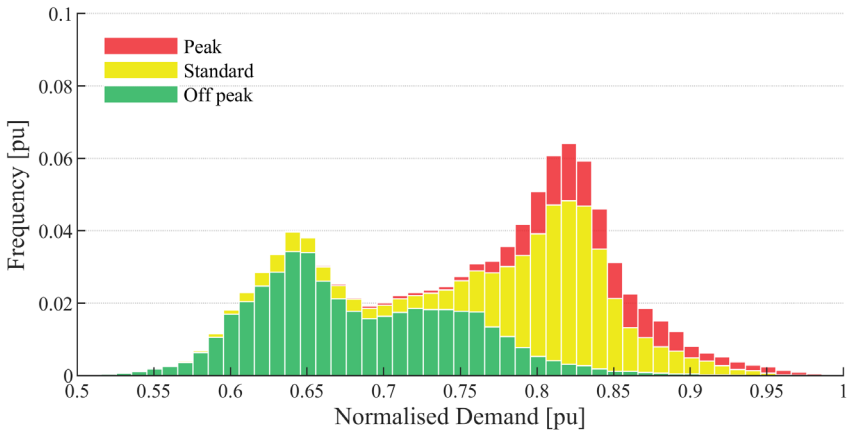


Figure 4.3: The detrended, normalised South African national demand profile histogram indicating the peak, standard and off-peak ToU periods from 2010 to 2014.

4.3.2 Baseline case studies

The results of the optimal spatial capacity allocation case studies are evaluated comparatively against four baseline case study scenarios of interest. These include the following:

Baseline case study 1: The generation capacities are assigned to the averaged temporal profiles of the individual REDZs such that the aggregated RE generation profile produces maximum cumulative energy yield.

Baseline case study 2: The wind and solar generation capacities are divided equally amongst the REDZs according to the RE technologies listed in Table 2.1. The wind generation capacity P_c^w is therefore divided equally between the Overberg, Komsberg, Cookhouse, Stormberg, Springbok and Beaufort West REDZs, while the solar generation capacity P_c^s is divided equally between all REDZs.

Baseline case study 3: The generation capacities are assigned to the clusters obtained by clustering the temporal profiles such that the aggregated RE generation profile produces maximum cumulative energy yield.

Baseline case study 4: The wind and solar generation capacities are divided equally amongst the wind and solar clusters obtained by clustering the ToU feature vectors. This baseline case study thus represents an unoptimised spatial allocation of the generation capacity, using the same clustered input profiles as for the optimisation case studies to follow.

Baseline case studies 1 and 3 are of interest because they are representative of current RE siting practices, whereby IPPs typically site the RE plants such that the cumulative annual energy production is maximised in order to take advantage of the flat feed-in RE tariff. This siting approach thus takes no cognisance of the temporal characteristics of the demand profile and the possibility of siting RE generation capacities to optimise the power balance along temporal timelines. These are best-case scenarios from an energy production perspective. However, the scenarios may be impractical for various reasons, including the fact that the capacities are typically allocated to one spatial region for wind and one spatial region for solar. If a selected region is geographically small, placing all of the RE resource capacity in such a region would be unrealisable. These scenarios also suffer from the disadvantage that large generation capacities can go offline due to unfavourable weather conditions occurring in the selected regions.

Baseline case studies 2 and 4 represent unoptimised, equal assignment scenarios used for comparative purposes. Baseline case 4 is significant since the same input profiles are used in the optimisation case studies, which provides a direct comparison scenario.

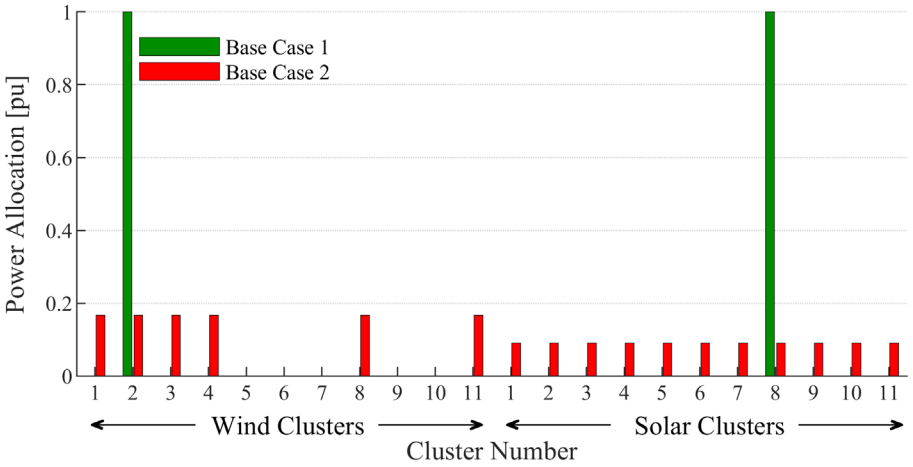
Table 4.2 summarises the baseline case studies considered in the investigation.

Table 4.2: Summary of baseline case studies.

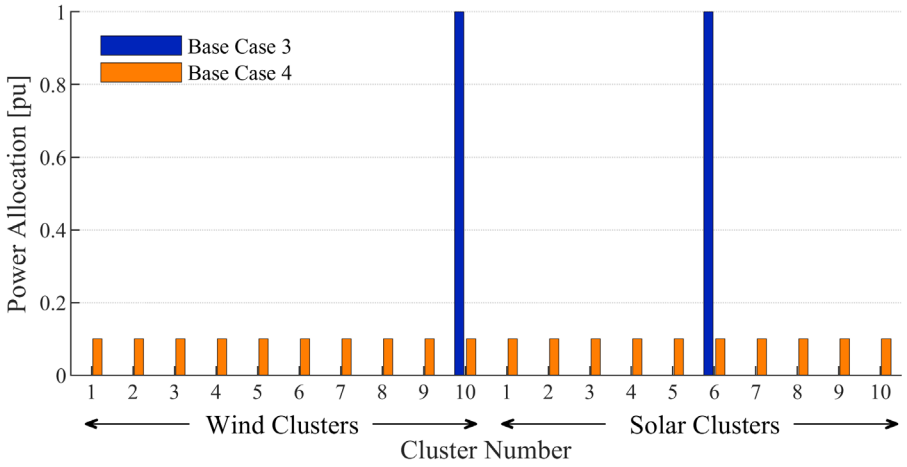
Baseline case study	Objective function	Input
1	Maximise cumulative energy yield	Averaged temporal profiles of the individual REDZs
2	Equal weight allocation to individual technology specific REDZs	Averaged temporal profiles of the individual REDZs
3	Maximise cumulative energy yield	Average profiles of individual temporal clusters
4	Equal weight allocation to individual clusters	Averaged profiles of the individual ToU clusters

Figure 4.4(a) depicts the normalised capacity allocations to the individual REDZs for baseline case studies 1 and 2. Figure 4.4(b) depicts the normalised capacity allocations to the temporal clusters and

ToU clusters for baseline case studies 3 and 4, respectively. As expected, by optimising the allocations for maximum cumulative energy yield, the capacities are assigned to a single entity for each of the wind and solar allocations.



(a) Capacity assignments to the individual Renewable Energy Development Zones



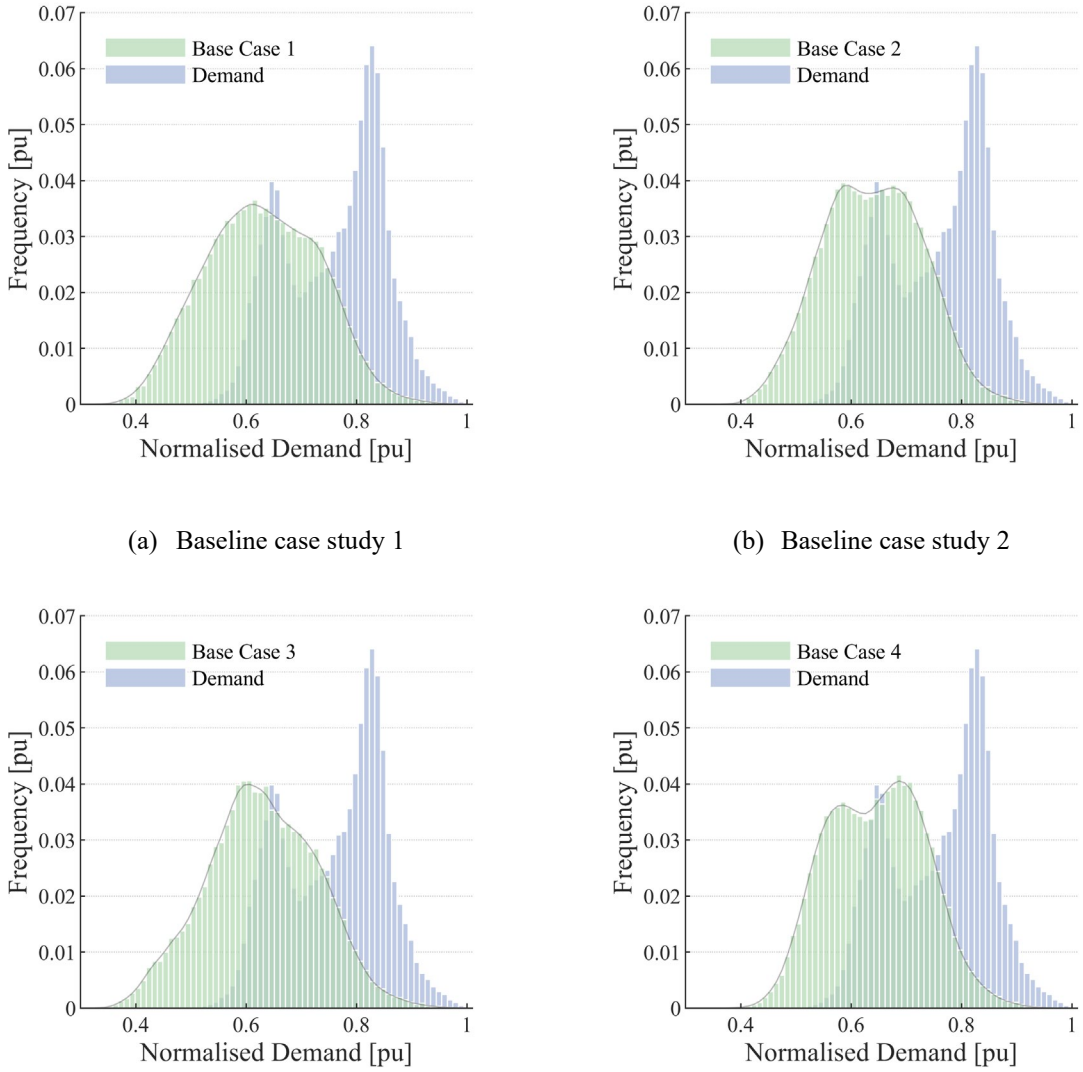
(b) Capacity allocations to the individual clusters

Figure 4.4: Normalised geospatial capacity allocations: (a) Baseline case studies 1 and 2, and (b) Baseline case studies 3 and 4.

Figure 4.5 shows the resultant normalised residual load profile histograms associated with each baseline case study, plotted against the original demand profile histogram to provide a visual comparison of the impacts. Baseline case study 1, shown in Figure 4.5(a), depicts a maximum frequency of occurance at approximately 0.59 pu power, whereas baseline case study 2, shown in Figure 4.5(b), depicts the highest frequency of occurance at approximately 0.58 pu, with a noticeable dip at 0.63 pu. In baseline case study 3, shown in Figure 4.5(c), the maximum frequency of occurance is at approximately 0.58 pu. Baseline case study 4, shown in Figure 4.5(d), depicts the highest frequency of occurance at approximately 0.7 pu, with another high frequency of occurance at approximately 0.57 pu, and a noticeable dip at 0.62 pu.

The histograms shown in Figure 4.5 are shown in Figure 4.6 as normalised probability density distributions, plotted against the normalised demand profile for comparison. All of the distributions

depict a significant reduction in the frequency of occurrence above 0.8 pu power compared to the normalised demand histogram. This study essentially examines to what extend further reduction in the frequency of occurrence of demand values above set thresholds can be achieved by optimised spatial capacity allocation.



(a) Baseline case study 1

(b) Baseline case study 2

(c) Baseline case study 3

(d) Baseline case study 4

Figure 4.5: Histograms of the residual load profiles for the various baseline case studies.

Figure 4.6 shows that both baseline case studies 1 and 3 depict a maximum frequency of occurance at approximately 0.6 pu power. These scenarios also service more energy at the lower load values, i.e., between 0.35 pu to 0.5 pu power, compared to baseline case studies 2 and 4, where more energy is serviced in the higher range, i.e., between the 0.65 pu to 0.75 pu power. This confirms that optimisation, even in the case of optimising cumulative energy yield as in scenarios 1 and 3, reduce the frequency of occurrence of higher load values compared to equal assignment unoptimised cases. However, it is noticeable that all of the baseline case studies perform very similar in the highest residual load range, i.e., above approximately 0.85 pu. Baseline case study 4 displays a notable dip in the frequency of occurrence at approximately 0.64 pu but seems to serve this energy at a higher load value, i.e., around 0.7 pu.

Overall, the results for the baseline case studies confirm that the geospatial capacity allocation strategy can have a major impact on the statistical distribution of the residual load profile.

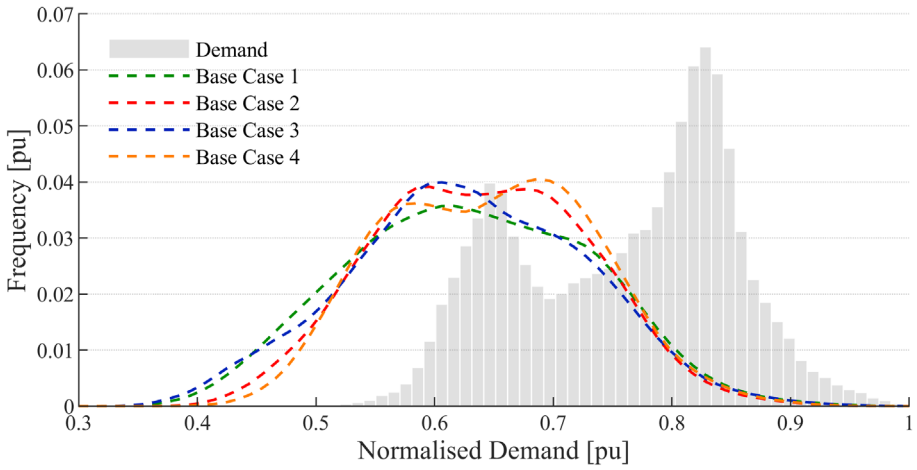


Figure 4.6: Probability density distributions of the residual load profiles for the various baseline case studies.

Table 4.3 provides the performance evaluation results for all baseline case studies. The normalised unweighted cumulative frequency of occurrence metric, F_1^T, represents the cumulative frequency of occurrence of residual power values above a set threshold 0.7 pu and 0.8 pu power, as a percentage of the total number occurrences, 43 824. The set threshold values of 0.7 pu and 0.8 pu power are used to determine the performance metric results for the baseline case studies, as to provide comparative results to quantify the impact of the optimisation case studies.

The normalised weighted cumulative frequency of occurrence metric, F_2^T, represents the cumulative energy above a set threshold, 0.7 pu and 0.8 pu power, as a percentage of the total energy.

F_3^T represents the frequency of occurrences above a given per unit range of above 0.9 pu of the maximum residual load value. These results remain unchanged dependent on the threshold value, this is because the baseline case studies are unoptimised and the capacity allocation weight vector remains unchanged between the weighted and unweighted scenarios.

Baseline case study 4 is shown to be the worst preforming case for a 0.7 pu threshold and baseline case 1 is shown to be the worst preforming case for a 0.8 pu threshold.

Table 4.3: Performance of the baseline case studies.

Baseline case study	F_1^T [%]	F_2^T [%]	F_3^T	F_1^T [%]	F_2^T [%]	F_3^T
	Unweighted			Weighted		
1: 0.7 pu	26.924	35.543	191	20.374	26.896	191
2: 0.7 pu	26.711	35.263	172	20.059	26.481	172
3: 0.7 pu	24.906	32.880	190	18.837	24.868	190
4: 0.7 pu	29.545	39.005	214	22.184	29.286	214
1: 0.8 pu	4.153	5.483	191	3.477	4.590	191
2: 0.8 pu	3.208	4.235	172	2.677	3.534	172
3: 0.8 pu	3.886	5.130	190	3.261	4.305	190
4: 0.8 pu	3.683	4.862	214	3.073	4.057	214

4.4 Optimisation case studies

The spatial capacity allocation framework is implemented for three optimisation case studies, which are evaluated comparatively against the four baseline case studies. These optimisation case studies include the following:

Optimisation case study 1: The weighted and the unweighted objective functions are implemented to derive optimised spatial capacity allocations for set thresholds of 0.7 pu and 0.8 pu power. This case study does not include capacity allocation constraints.

Optimisation case study 2: The weighted and the unweighted objective functions are implemented to derive optimised spatial capacity allocations for set thresholds of 0.7 pu and 0.8 pu power. This case study includes minimum capacity constraints, which represent the existing wind and solar plants within each cluster. Figure 4.7 depicts the existing wind and solar plants within South Africa where the encircled RE plants are included in the minimum constraints.

Optimisation case study 3: The unweighted objective function is implemented to derive optimised spatial capacity allocations for a set threshold of 0.7 pu power. This case study includes minimum and maximum capacity constraints, where the minimum capacity constraints are defined as in optimisation case study 2. For this case study scenario, the maximum constraints are derived from the recent GCCA study [181]. Figure 4.8 depicts the remaining generation capacity that can be evacuated by existing grid infrastructure for the various supply areas. The Northern Cape has no further supply capacity, besides existing and committed capacity in this region. The supply area capacity within the six areas accumulates approximately 10.5 GW of generation capacity. These generation limits are used to formulate the maximum constraints.

Table 4.4 summarises the optimisation case studies considered in the investigation.

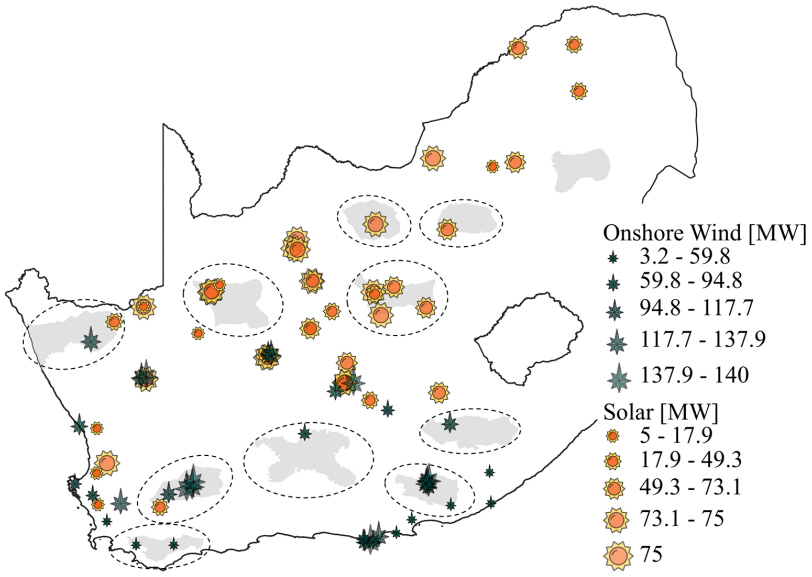


Figure 4.7: Existing renewable energy plants. Renewable energy plants included in the minimum optimisation constraints are encircled.

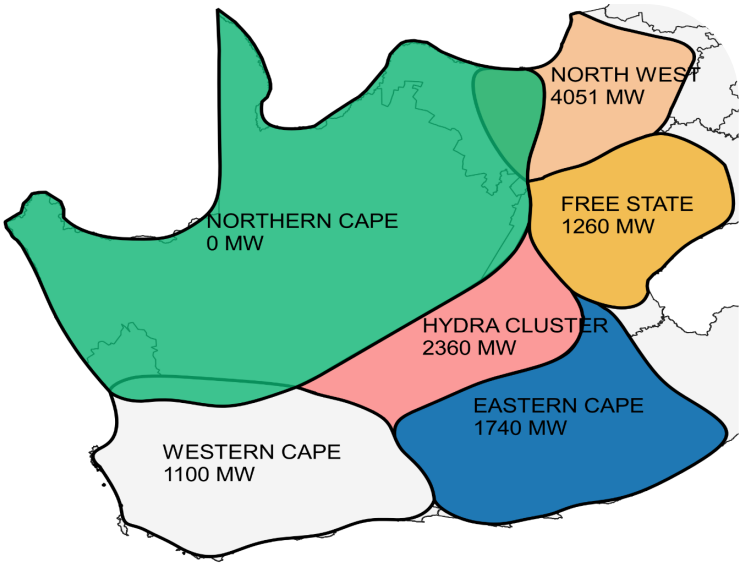


Figure 4.8: Power evacuation capacities for the South African supply areas [181].

Table 4.4: Summary of optimisation case studies.

Optimisation case studies		Objective Functions	Constraints
1	1.1	Minimise residual load events above thresholds $P_T = 0.7$ pu and $P_T = 0.8$ pu , unweighted objective function	None
	1.2	Minimise residual load events above thresholds $P_T = 0.7$ pu and $P_T = 0.8$ pu , unweighted objective function	None
2	2.1	Minimise residual load energy above thresholds $P_T = 0.7$ pu and $P_T = 0.8$ pu , unweighted objective function	Minimum
	2.2	Minimise residual load energy above thresholds $P_T = 0.7$ pu and $P_T = 0.8$ pu , unweighted objective function	Minimum
3		Minimise residual load energy above threshold $P_T = 0.7$ pu, unweighted objective function	Minimum and maximum

4.4.1 Optimisation case studies 1.1 and 1.2

The unconstrainted optimisation case studies 1.1 and 1.2 determine the spatial capacity allocations for optimal power balance without considering capacity and grid infrastructure limitations. The optimisation framework is implemented using the averaged profiles of the clusters obtained with ToU feature-based clustering. The unweighted and weighted objective functions defined by Equations (4.25) and (4.26) are implemented, using thresholds of 0.7 pu and 0.8 pu power.

Figure 4.9 shows the optimal capacity allocations obtained for optimisation case studies 1.1 and 1.2. The available wind generation capacity is allocated predominantly to clusters 3, 5 and 10 for both power thresholds in both case studies. The clustered wind resource map shown in Figure 3.10 depicts the geographic distribution of the clusters. Cluster 10 is represented in 6 of the 11 REDZ, but is located predominantly in Komsberg, Beaufort West and Stormberg. Although a high percentage of the capacity is allocated to cluster 10, the cluster represents a diverse geographic spread, which allows for spatial dispersion across REDZs. Figure 3.10 shows that cluster 3 is represented in the same 6 REDZs as cluster 10, but is located predominantly in the Springbok, Overberg and Komsberg REDZs. Cluster 5 is situated

in the northern region, predominately in the Springbok, Upington and Vryburg REDZs. This is of interest since the gazetted REDZs define the Upington and Vryburg REDZs as a solar integration region.

Figure 3.14(a) and Figure 3.14(b) show that the yield associated with cluster 3 occurs predominantly in the late afternoon period. In Figure 3.14, cluster 5 shows a dip in midday production, with increased yield in the morning and evening periods. This shows good complementary characteristics in comparison with the yield of the solar power produced during midday. Clusters 1, 4, 6, 7, 8 and 9 received little to no wind capacity allocation. Cluster 1, 4 and 8 are predominately situated in the north-eastern region, i.e., in the Kimberly, Emalahleni and Klerksdorp REDZs. Clusters 4 and 7 are situated in the southern and western regions but show a lower average yield in comparison with clusters 2, 3 and 10 located in the same region. This suggests that the lower wind speed locations within these regions are clustered together. By comparing clusters 4 and 7 with the underlying wind speed characteristics, shown in Figure 3.12, the wind speed map confirms low wind speeds within these clusters. Cluster 9 shown in Figure 3.14(a) depicts a high yield in the evening peak period, with a low yield in the morning peak period. Furthermore, the results show that the choice of objective function, i.e., weighted versus unweighted, as well as the threshold values, i.e., 0.7 pu or 0.8 pu, impact quite significantly on the allocations. The weighted objective function combined with the higher threshold value generally yields more diverse weight allocations to the clusters.

The available solar generation capacity is allocated to all solar clusters. For a threshold value of 0.7 pu, a large share is assigned to cluster 9 for both objective functions. For a threshold value of 0.8 pu, the largest share is assigned to cluster 1 for both objective functions. For the unweighted objective function with a threshold value of 0.8 pu, a fair share is also assigned to cluster 6. Figure 3.24 displays the ToU feature-based solar clusters, where clusters 1 and 9 are located in the north-western region, and cluster 6 is located in the north-eastern region. Cluster 6 shown in Figure 3.27(b) depicts a relatively high yield in the morning and midday for the high demand season. Clusters 1 and 9 shown in Figure 3.27(b) and Figure 3.27(c) depict a relatively high yield during the midday and towards the evening peak periods in the low and high demand season.

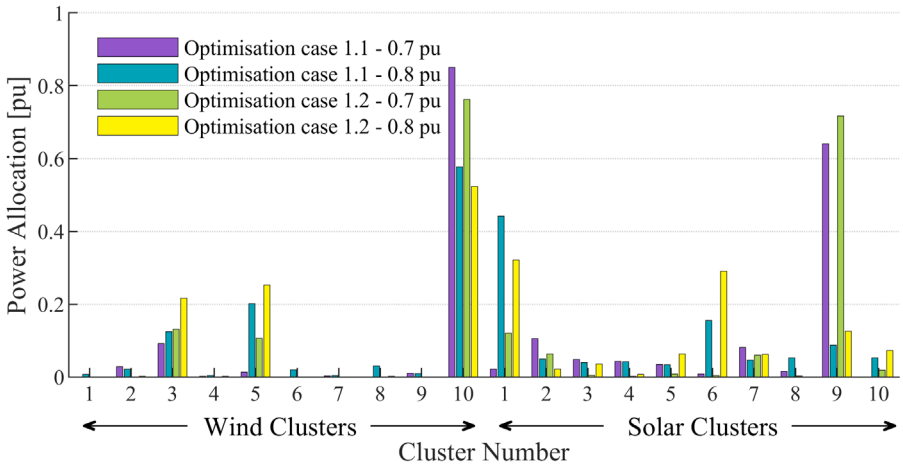
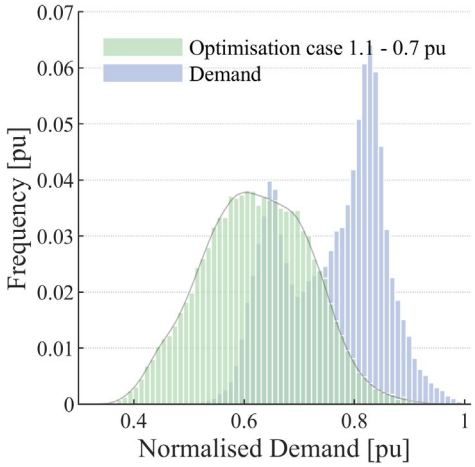


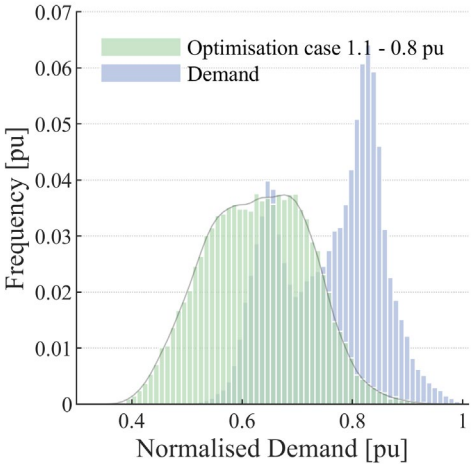
Figure 4.9: Normalised geospatial capacity allocations for optimisation case studies 1.1 and 1.2, for thresholds of 0.7 pu and 0.8 pu power.

Figure 4.10 depicts the histograms of the normalised residual load profiles obtained for optimisation case studies 1.1 and 1.2, for set thresholds of 0.7 pu and 0.8 pu power. The histogram of the original demand profile is shown to provide a basis for visual comparison.

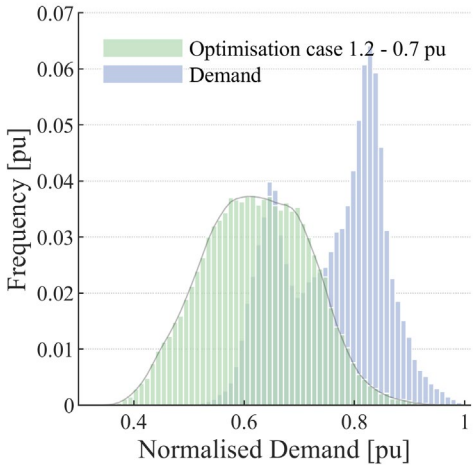
Both the unweighted and weighted objective functions, for a threshold of 0.7 pu, exhibit a maximum frequency of occurrence at approximately 0.58 pu power. For a threshold of 0.8 pu, the maximum frequency of occurrence is shifted to approximately 0.72 pu power for both objective functions.



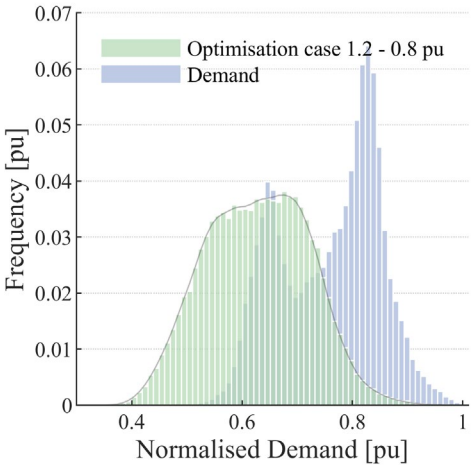
(a) Unweighted objective function with a threshold value of 0.7 pu



(b) Unweighted objective function with a threshold value of 0.8 pu



(c) Weighted objective function with a threshold value of 0.7 pu



(d) Weighted objective function with a threshold value of 0.8 pu

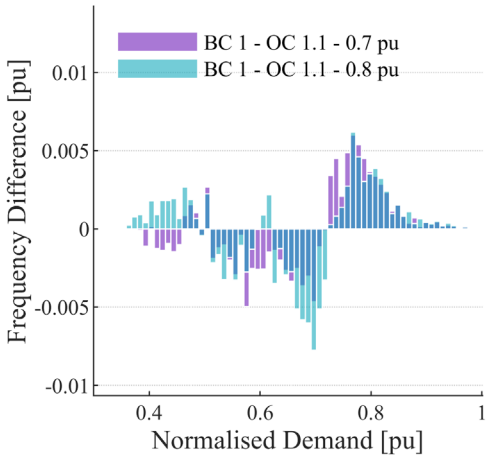
Figure 4.10: Histograms of the optimised residual load profile for optimisation case studies 1.1 and 1.2.

In order to evaluate the performance of the optimised capacity allocation, the residual load histograms for the optimisation case studies are subtracted from each of the baseline case study histograms. This yields bar charts that display the differences in frequency of occurrences between the optimisation case study results and the baseline case studies. Therefore, a positive value shows that the baseline case study has a lower frequency of occurrence for that normalised power range, while a negative value shows that the optimisation case study has a higher frequency of occurrence for that power range. Figure 4.11 and Figure 4.12 depict bar charts of the difference in frequency of occurrence for each comparative scenario.

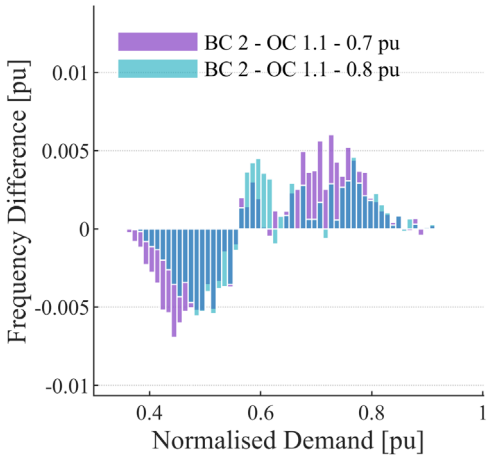
In Figure 4.11 and Figure 4.12, all instances display predominately positive differences in the range above 0.78 pu. This indicates that the optimisation framework successfully reduced the frequency of occurrence above the set thresholds when compared to all baseline case studies.

Baseline case studies 1 and 3 represent the capacity allocation for maximum cumulative yield, where it is notable that the optimised placement achieves reduction in the frequency of occurrence above the set threshold values.

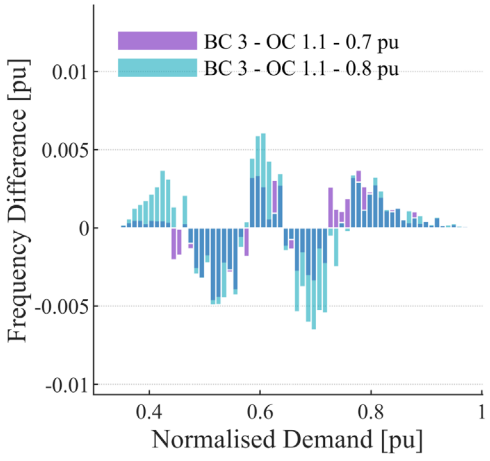
The unweighted cases shown in Figure 4.11 show a distinct difference between the 0.7 pu and 0.8 pu thresholds, where the 0.7 pu threshold case shows a higher reduction in hourly occurrences closer to the set threshold and the 0.8 pu threshold case shows a higher reduction in hourly occurrences in the higher power range, closer to 1 pu. Figure 4.11(b) and Figure 4.11(d) depict a higher reduction in hourly occurrences when compared to Figure 4.11(a) and Figure 4.11(c), which is accounted to the difference in the baseline case function objectives, summarised in Table 4.2, which translates to the differences in the probability density function plots shown in Figure 4.6.Figure 4.12 depicts the weighted optimisation case study, where a greater importance is placed on higher power values, closer to 1 pu. Comparing each diagram in Figure 4.11 against the same case diagrams in Figure 4.12, although marginal, all instances depict a shift towards 1 pu. Overall, the optimisation strategy clearly leverages the temporospatial characteristics of the resource profiles to achieve the desired outcomes.



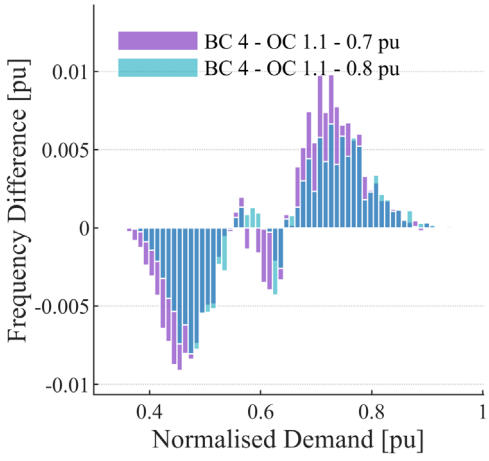
(a) Baseline case study 1: Maximise energy yield for assignment to REDZs profiles



(b) Baseline case study 2: Equal assignment to averaged REDZs profiles

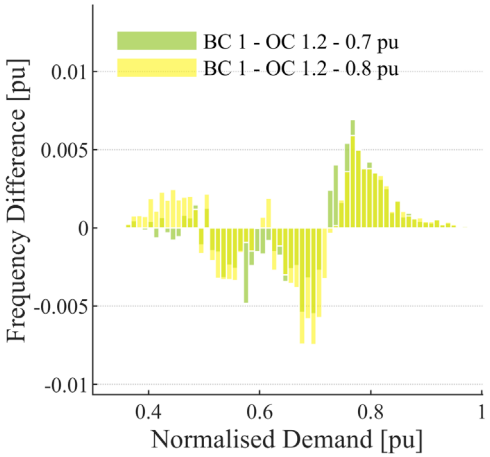


(c) Baseline case study 3: Maximise energy yield for assignment to temporal clusters

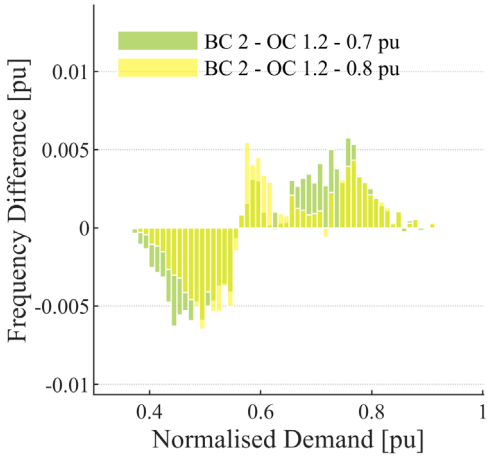


(d) Baseline case study 4: Equal assignment to ToU cluster profiles

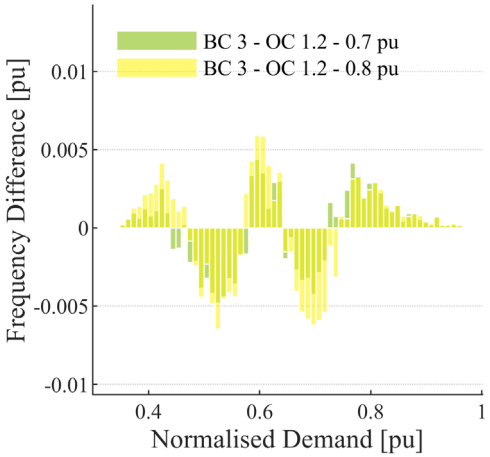
Figure 4.11: Difference in the histograms of unweighted optimisation case study 1.1 and the baseline case studies.



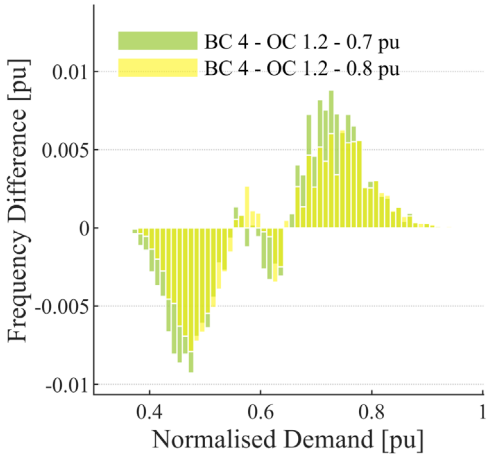
(a) Baseline case study 1: Maximise energy yield for assignment to averaged REDZs profiles



(b) Baseline case study 2: Equal assignment to averaged REDZs profiles



(c) Baseline case study 3: Maximise energy yield for assignment to temporal clusters



(d) Baseline case study 4: Equal assignment to ToU cluster profiles

Figure 4.12: Difference in the histograms of weighted cases study 1.2 and the baseline case studies.

Table 4.5 provides the performance evaluation results for the optimisation case studies. The normalised unweighted cumulative frequency of occurrence metric, F_1^T, represents the cumulative frequency of occurrence of residual power values above a set threshold 0.7 pu and 0.8 pu power, as a percentage of the total number occurrences, 43 824. The unweighted scenarios, for both thresholds, show a higher percentage of occurrence compared to the weighted cases.

The normalised weighted cumulative frequency of occurrence metric, F_2^T, represents the cumulative energy above a set threshold, 0.7 pu and 0.8 pu power, as a percentage of the total energy. The unweighted scenarios, for both thresholds, also show higher percentages of occurrence compared to the weighted cases.

F_3^T represents the frequency of occurrences above a given per unit range of 0.9 pu of the maximum residual load value. For the 0.7 pu threshold, the weighted case shows a reduction in occurrences, while the 0.8 pu threshold shows an increase in occurrences above 90 % of the maximum residual load value.

Table 4.6 lists the differences between the baseline case study results and the optimisation case study results. The difference in performance metrics is calculated by subtracting the optimisation case study

performance results, shown in Table 4.5, from the baseline case study performance results, shown in Table 4.3. Therefore, if the results are positive then the optimisation case studies out preformed the baseline case studies and if the results are negative then the baseline case studies out preformed the optimisation case studies. Table 4.6 shows that all values are positive, which indicates that the optimisation case study 1 outperformed all baseline case scenarios. The greatest improvement is seen between baseline case study 4 and the optimisation case study for a threshold of 0.7 pu power. For the unweighted case, there is a 7.18 % improvement in the frequency of occurrence of residual power values and a 9.48 % improvement in the cumulative energy yield when compared to the baseline case study 4 scenario. The frequency of occurrences of residual power values above 90 % of the maximum residual load value is reduced by 78 instances. This is a significant result since the baseline case study 4 represents the unoptimised case for the same statistical input clusters. Similarly for the weighted case, there is a 5.43 % improvement in the frequency of occurrence of residual power values and a 7.16 % improvement in the cumulative energy yield when compared to the baseline case study 4 for this set threshold value. The frequency of occurrences of residual power values above 90 % of the maximum residual load values is reduced by 84 instances. This increase from 78 to 84 reduced instances shows that the weighted optimisation objective succeeded in reducing more instances of occurrence for higher residual load values. This is not the case for the 0.8 pu threshold scenarios, which could be because the optimisation range is shorter, and the reduction in the higher residual load values is seen closer to the maximum residual load value, i.e., above 95 %.

Overall, the optimisation case studies all outperformed the baseline case studies for load instances above the set threshold, with improvements ranging from 1 % to 9.5 % dependent on the objective function. This shows a significant result since these high residual load values quantify the most expensive operating time when replaced by ancillary services.

Table 4.5: Performance of the optimisation case studies 1.1 and 1.2.

Optimisation case study	F_1^T [%]	F_2^T [%]	F_3^T	F_1^T [%]	F_2^T [%]	F_3^T
	0.7 pu			0.8 pu		
1.1	22.278	29.410	136	2.515	3.320	136
1.2	16.756	22.120	129	2.056	2.714	139

Table 4.6: Difference between performance results between the baseline case studies and the optimisation case studies.

	F_1^T [%]	F_2^T [%]	F_3^T	F_1^T [%]	F_2^T [%]	F_3^T
	Unweighted			Weighted		
1: 0.7 pu	4.561	6.022	55	3.614	4.771	61
2: 0.7 pu	4.349	5.742	36	3.300	4.356	42
3: 0.7 pu	2.544	3.359	54	2.078	2.744	60
4: 0.7 pu	7.183	9.483	78	5.425	7.161	84
1: 0.8 pu	1.625	2.145	58	1.414	1.866	54
2: 0.8 pu	0.680	0.898	39	0.614	0.810	35
3: 0.8 pu	1.358	1.792	57	1.197	1.581	53
4: 0.8 pu	1.155	1.524	81	1.010	1.333	77

Figure 4.13 and Figure 4.14 depict the clustered wind and solar resource maps together with the optimised allocation capacity for the 0.8 pu power threshold for the unweighted optimisation case study 1.1. This optimisation case is chosen as an example to depict the cluster formations together with the capacity allocation values, to form useful RE geospatial capacity allocation siting maps.

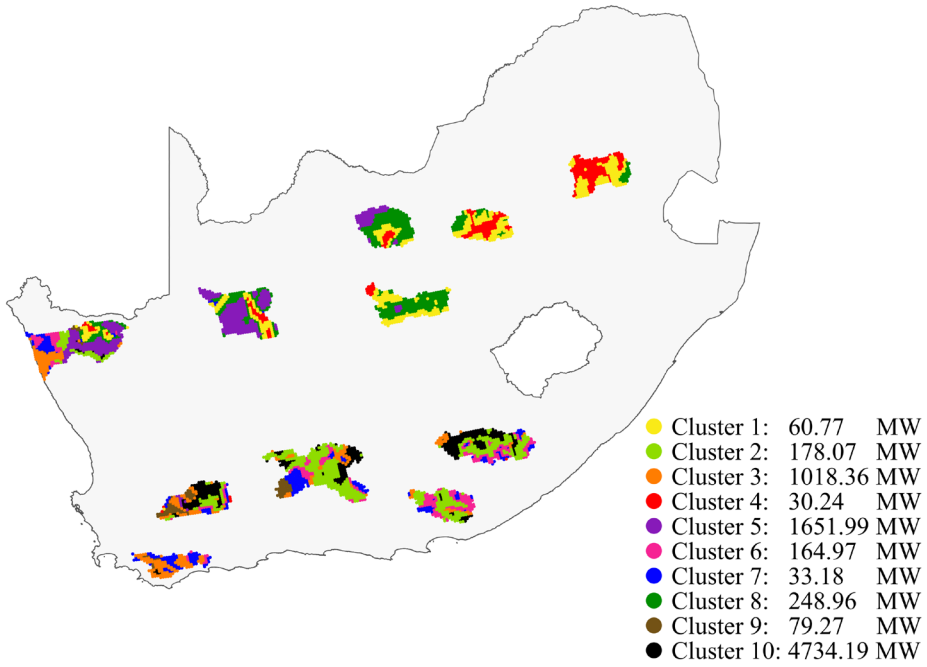


Figure 4.13: Clustered wind resource map together with the optimised capacity allocations for optimisation case study 1.1, for a 0.8 pu threshold.

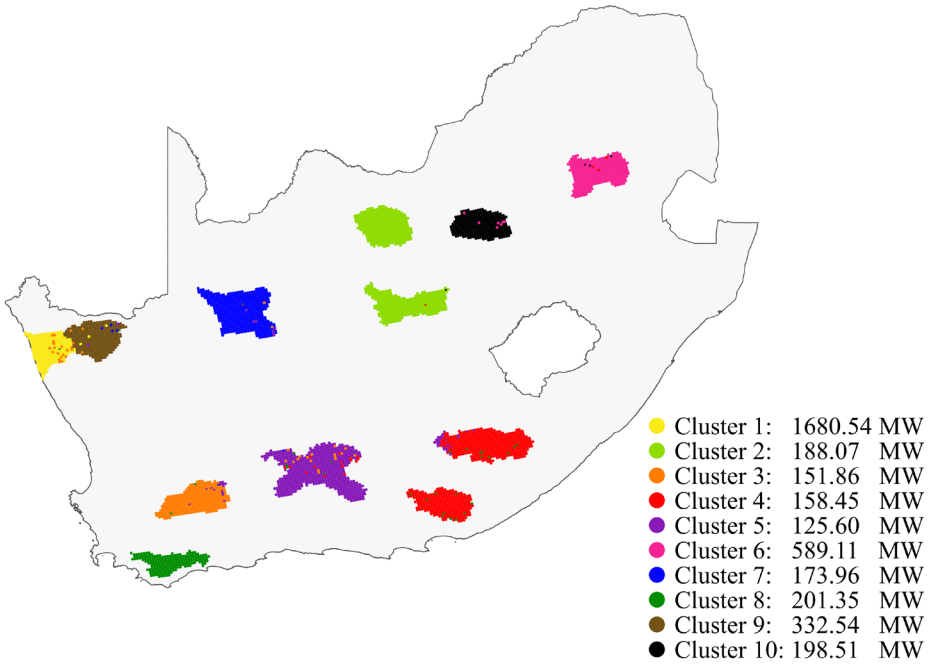


Figure 4.14: Clustered solar resource map together with the optimised capacity allocations for optimisation case study 1.1, for a 0.8 pu threshold.

4.4.2 Optimisation case studies 2.1 and 2.2

The constrained optimisation case studies 2.1 and 2.2 determine the spatial capacity allocation scenarios for optimal power balance, while considering minimum capacity constraints defined by the existing wind and solar generation units within the clustered regions. The optimisation framework is implemented using the averaged profiles of the clusters obtained with the ToU feature-based clustering approach. The unweighted and weighted objective functions defined by Equations (4.25) and (4.26) are implemented, using thresholds of 0.7 pu and 0.8 pu power.

Figure 4.15 shows the optimal capacity allocations obtained for optimisation case studies 2.1 and 2.2. The minimum capacity constraints are indicated in Figure 4.15 by transparent grey bars plotted over the optimal capacity assignments. The available wind generation capacity is allocated predominantly to cluster 10 for both power thresholds in both case studies. All capacity is allocated to cluster 10, apart from the capacity committed to the minimum constraints for optimisation case study 2.1 for a threshold of 0.7 pu. Optimisation case study 2.1 for a threshold of 0.8 pu, allocates majority of additional capacity to clusters 2, 3 and 5. Optimisation case study 2.2 for a threshold of 0.8 pu, allocates majority of the additional capacity to clusters 3, 5, 7 and 9. In Figure 3.10, cluster 2, 3, 7, 9 and 10 are situated in the southern region, specifically in Overberg, Komsberg, Beaufort West, Cookhouse and Stormberg. High wind speed patterns are depicted within these areas, which is confirmed in Figure 3.14(a), specifically in the evening peak period. Cluster 5 is situated in the northern region, predominately in the Springbok, Upington and Vryburg REDZs. This is of interest since the gazetted REDZs define the Upington and Vryburg REDZs as a solar integration region. In Figure 3.14(a)-(c) it is noted that cluster 2, 3, 7, 9 and 10 display similar annual daily yield profiles, with a peak in the evening peak period. In Figure 3.14, cluster 5 shows a dip in midday production, with increased yield in the morning and evening periods. This shows good complementary characteristics in comparison with the yield of the solar power produced during midday. Clusters 1, 4, 6, 7 and 8 received little to no capacity allocation. Clusters 1, 4 and 8 are predominately situated in Emalahleni, Vryburg, Kimberly and Klerksdorp in the northeast region. Figure 3.12 compares the clusters with the underlying wind speed characteristics, where the regions depict low wind speed characteristics. Comparing clusters 6 and 7 with the underlying wind speed characteristics, shown in Figure 3.12, these clusters, overlayed onto the wind speed map, also show low wind speed characteristics.

The available solar generation capacity is allocated to all solar clusters. For a threshold value of 0.7 pu, a large share is assigned to cluster 9 and for a threshold value of 0.8 pu, a large share is assigned to cluster 10. Figure 3.24 shows that clusters 9 and 10 are situated in Springbok and Klerksdorp, respectively. Figure 3.27 shows that cluster 9 has the highest yield during the evening peak periods, and cluster 10 shows high yield in the morning period. The cluster with the lowest assigned capacity is cluster 4, which is situated in Stormberg and Cookhouse. Comparing Figure 2.3 and Figure 3.24, it shows that cluster 4 has of the lowest average GHI values for that region.

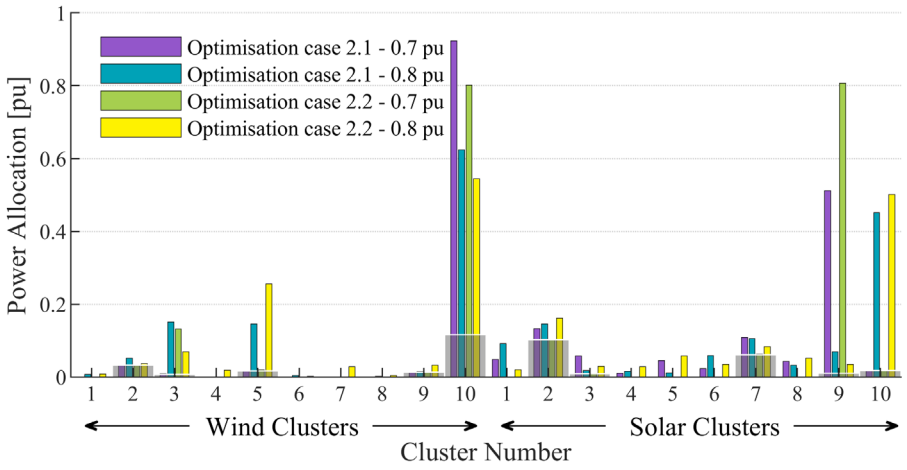
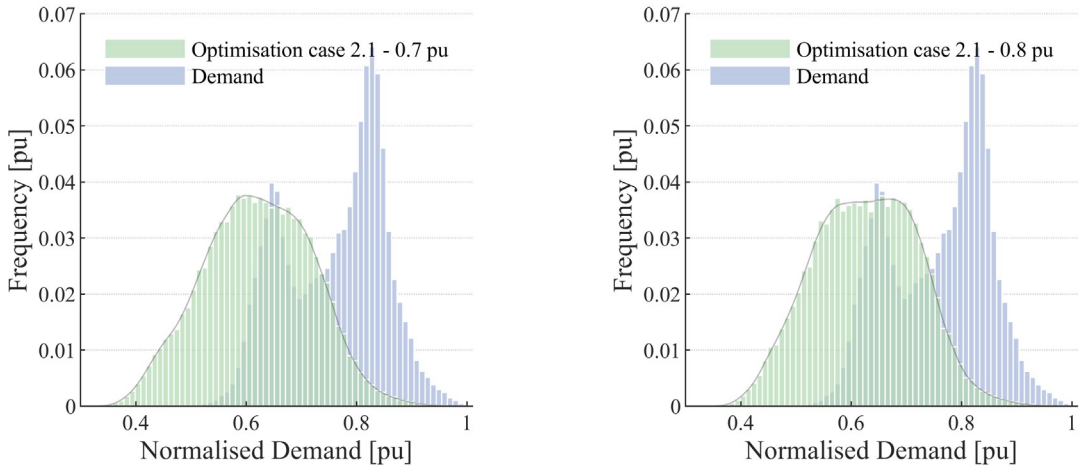


Figure 4.15: Normalised geospatial capacity allocations for optimisation case studies 2.1 and 2.2, for set thresholds of 0.7 pu and 0.8 pu power.

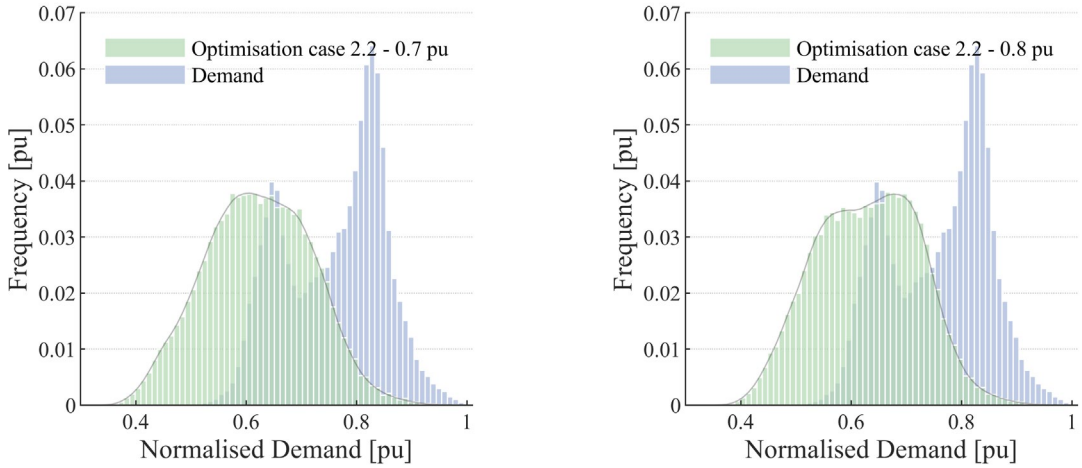
Figure 4.16 depicts the histograms of the normalised residual load profiles obtained with optimisation case studies 2.1 and 2.2, for set thresholds of 0.7 pu and 0.8 pu power. The histogram of the original demand profile is shown to provide a basis for visual comparison.

Optimisation case 2.1 depicts a maximum frequency of occurance at approximately 0.58 pu power for a threshold of 0.7 pu, whereas for a threshold of 0.8 pu, the maximum frequency of occurance is shifted to approximately 0.71 pu power. Optimisation case 2.2 shows the highest frequency of occurance at approximately 0.58 pu for a threshold of 0.7 pu, whereas for a threshold of 0.8 pu, the maximum frequency of occurance is shifted to approximately 0.71 pu power.



(a) Unweighted objective function with a threshold value of 0.7 pu

(b) Unweighted objective function with a threshold value of 0.8 pu



(c) Weighted objective function with a threshold value of 0.7 pu

(d) Weighted objective function with a threshold value of 0.8 pu

Figure 4.16: Histograms of the optimised residual load profiles for optimisation case studies 2.1 and 2.2

To evaluate the performance of the optimised capacity allocation, the residual load histograms for the optimisation case studies are subtracted from each of the baseline case study histograms. This yields bar charts that display the differences in frequency of occurrences between the optimisation case study results and the baseline case studies.

Figure 4.17 and Figure 4.18 depict bar charts of the difference in the frequency of occurrence for residual load instances above a set threshold, for each comparative scenario.

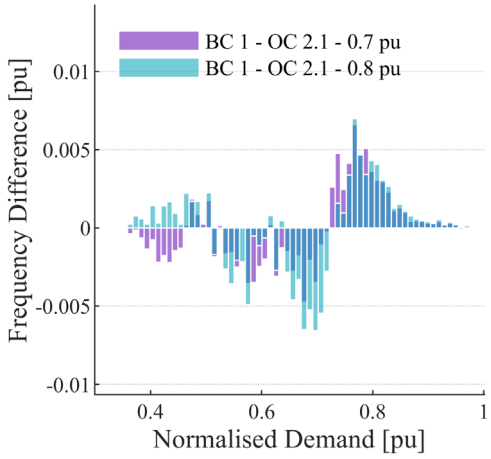
In Figure 4.17(a) and (c) and Figure 4.18(a) and (c) all of the instances display positive differences in the range above approximately 0.75 pu. In Figure 4.17(b) and (d) and Figure 4.18(b) and (d) all of the instances display positive differences in the range above approximately 0.68 pu. Figure 4.17(d) and Figure 4.18(d) represent baseline case study 4, which is an unoptimised case which uses the same input profile as the optimisation case studies. These profiles show a large decrease in the frequency of occurrence of the residual load values for the 0.7 pu set threshold. This indicates that optimising the geospatial capacity allocation, rather than allocating equally between clusters, successfully aids in reducing the frequency of occurrence of residual load power values above the set thresholds.

Overall, the results indicate that the optimisation framework successfully reduces the frequency of occurrence of residual load power values above the set thresholds when compared to all baseline case studies.

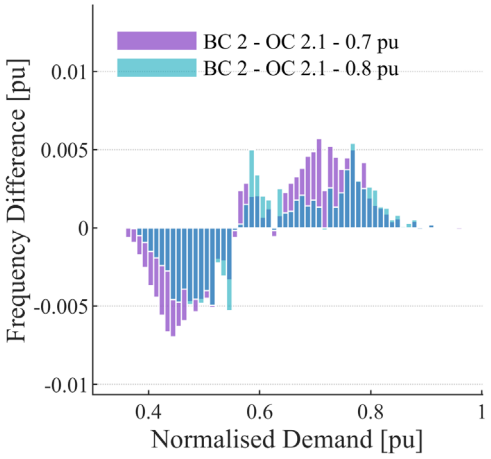
The unweighted cases shown in Figure 4.17 show a distinct difference between the 0.7 pu and 0.8 pu thresholds, where the 0.7 pu threshold case shows a higher reduction in hourly occurrences closer to the set threshold and the 0.8 pu threshold case shows a higher reduction in hourly occurrences in the higher power range, closer to 1 pu. Figure 4.17(d) depicts a higher reduction in hourly occurrences when compared to Figure 4.17(a)-(c).

Figure 4.18 depicts the weighted optimisation case study, where a greater importance is placed on higher power values, closer to 1 pu. Comparing each diagram in Figure 4.17 against the same case diagrams in Figure 4.18, although marginal, all instances depict a shift towards 1 pu.

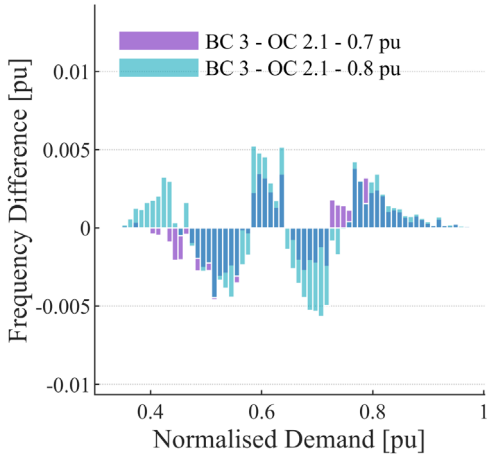
Overall, the optimisation strategy clearly leverages the temporospatial characteristics of the resource profiles to achieve the desired outcomes. The results show similar capacity allocations as optimisation case 1, apart from the minimum constraint allocations included in the capacity weight vectors.



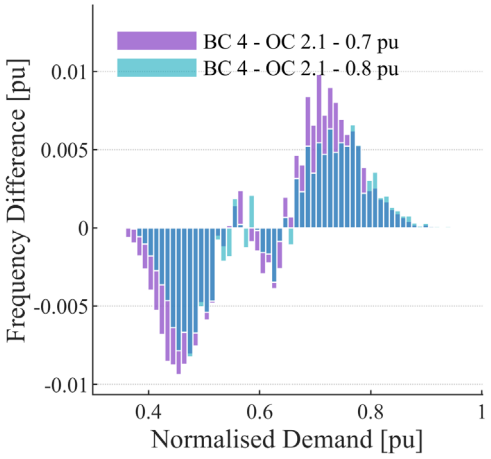
(a) Baseline case study 1: Maximise energy yield for assignment to REDZs profiles



(b) Baseline case study 2: Equal assignment to averaged REDZs profiles

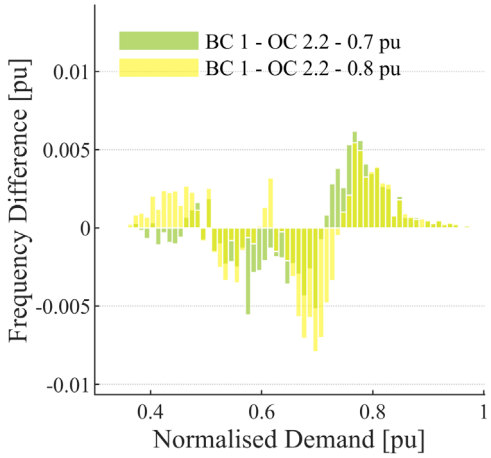


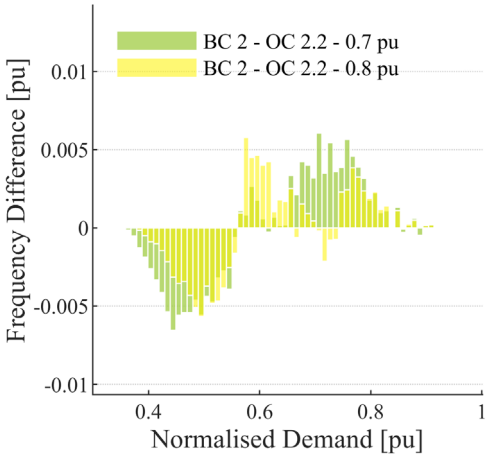
(c) Baseline case study 3: Maximise energy yield for assignment to temporal clusters



(d) Baseline case study 4: Equal assignment to ToU cluster profiles

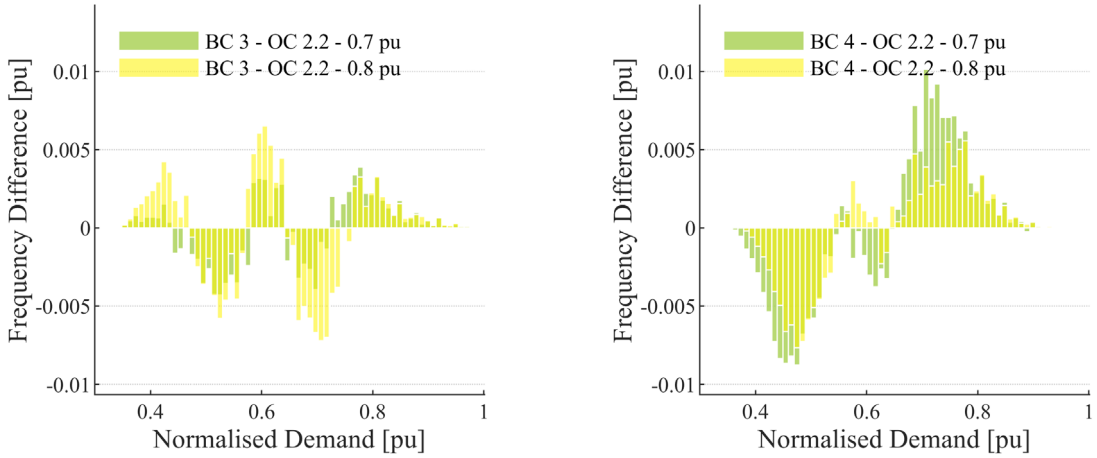
Figure 4.17: Difference in the histograms of unweighted optimisation case study 2.1 and the baseline case studies.





(a) Baseline case study 1: Maximise energy yield for assignment to averaged REDZs profiles

(b) Baseline case study 2: Equal assignment to averaged REDZs profiles



(c) Baseline case study 3: Maximise energy yield for assignment to temporal clusters

(d) Baseline case study 4: Equal assignment to ToU cluster profiles

Figure 4.18: Difference in the histograms of weighted cases study 2.2 and the baseline case studies.

Table 4.7 provides the performance evaluation results for the optimisation case studies 2.1 and 2.2. Table 4.8 lists the differences between the baseline case study results and the optimisation case study results. Table 4.8 shows that all values are positive, which indicates that the optimisation case study 2 outperformed all baseline case studies. The greatest improvement is seen between baseline case study 4 and the optimisation case study for a threshold of 0.7 pu power. For the unweighted case, there is a 7.05 % improvement in the frequency of occurrence of residual power values above the set threshold and a 9.3 % improvement in the cumulative energy yield above the set threshold when compared to baseline case study 4. The difference in the frequency of occurrences of residual power values above 90 % of the maximum residual load value, between the optimisation case 2 and baseline case 4 study shows a reduction of 84 instances. These residual load values quantify the most expensive operating time when replaced by ancillary services. This is a significant result since the baseline case study 4 represents the unoptimised case for the same statistical input clusters. Similarly for the weighted case, there is a 5.3 % improvement in the frequency of occurrence of residual power values and a 7 % improvement in the cumulative energy yield when compared to the baseline case study 4, for a set threshold of 0.8 pu power. The difference in the frequency of occurrences of residual power values above 90 % of the maximum residual load value, between the optimisation case 1 and baseline case 4 shows a reduction of 85 instances.

The optimisation case study 2 shows similar results to the optimisation case study 1, which is to be expected since the inclusion of the minimum capacity constraints did not have a large impact on the capacity weight vector allocations.

Table 4.7: Performance of the optimisation case studies 2.1 and 2.2.

Optimisation case study	F_1^T [%]	F_2^T [%]	F_3^T	F_1^T [%]	F_2^T [%]	F_3^T
	0.7 pu			0.8 pu		
2.1	22.396	29.567	130	2.478	3.271	141
2.2	16.603	21.918	136	2.108	2.783	138

Table 4.8: Difference between performance results between the baseline case studies and the optimisation case studies.

	F_1^T [%]	F_2^T [%]	F_3^T	F_1^T [%]	F_2^T [%]	F_3^T
	Unweighted			Weighted		
1: 0.7 pu	4.429	5.847	61	3.494	4.613	62
2: 0.7 pu	4.217	5.567	42	3.180	4.198	43
3: 0.7 pu	2.412	3.184	60	1.958	2.585	61
4: 0.7 pu	7.051	9.308	84	5.305	7.003	85
1: 0.8 pu	1.622	2.142	56	1.357	1.791	54
2: 0.8 pu	0.678	0.895	37	0.557	0.735	35
3: 0.8 pu	1.355	1.789	55	1.140	1.505	53
4: 0.8 pu	1.152	1.521	79	0.953	1.258	77

4.4.3 Optimisation case study 3

Constrained optimisation case study 3 determines the spatial capacity allocation for optimal power balance while considering minimum and maximum capacity constraints. The minimum constraint is defined by the existing wind and solar generation units within the clustered regions. The maximum constraints are set according to the capacity limits shown in Figure 4.8.

The optimisation framework is implemented using the average ToU feature-based clusters as input. The resultant residual load histogram is targeted, where unweighted occurrence above a set threshold of 0.7 pu power is minimised. The maximum constraints are derived for grouped areas. The grouped areas are derived from Figure 4.8 and are defined in Table 4.9.

Table 4.9: Capacity supply constraints per zone, depicted in Figure 4.8.

Region		REDZs within the Region	Capacity Supply Limit
1	Western cape	Overberg, Komsberg, Beaufort West	1.1 GW
2	Eastern cape	Cookhouse, Stormberg, Beaufort West	1.74 GW
3	Free state	Klerksdorp, Kimberly	1.26 GW
4	Northwest	Klerksdorp	4.051 GW
5	Northern cape	Springbok, Upington, Vryburg, Kimberly	0 GW
6	Unconstraint region	Emalahleni	None

The combined capacity allocation for wind and solar, defined by the IRP2019, amounts to 12 GW, 8.2 GW of wind and 3.8 GW of solar [14]. The minimum capacity allocation is defined by the existing RE generation plants within the REDZs, this summates to 1.526 GW of wind capacity and 0.769 GW of solar capacity. The remaining capacity to be allocated amounts to 6.657 GW of wind capacity and 3.05 GW of solar capacity. The supply area is divided into 6 regions, listed in Table 4.9, where regions 1 to 5 have a set maximum constraint, as shown in Figure 4.8. The Hydra cluster is excluded since no REDZs fall within this area. The 6th region is defined as all areas outside of the constrained regions, where this region has no maximum constraint set. This is a vital region, since the summation of the maximum

constraints for regions 1 to 5 equals to 8.151 GW, which this is less than the power capacity to be allocated, namely, 6.657 GW of wind capacity and 3.05 GW of solar capacity. Emalahleni is the only REDZs situated in the unconstraint region, therefore all excess capacity must be allocated to this REDZs. Regions 1 and 2 are situated in the south of the country, where Figure 2.2 depicts the favourable wind speed characteristics. Regions 3 and 4 are situated in the northeast part of the country, where Figure 2.3 depicts favourable solar GHI characteristics. Regions 1 and 2 are only able to accommodate 2.84 GW, which implies that the remaining wind capacity will be forced into the southern part of the country. Region 5 has a maximum constraint set to zero, therefore no additional capacity can be allocated to this region. The existing capacity is constraint per cluster and added to the final capacity allocation after the maximum constraint limits have been taken into consideration. The maximum supply constraints define the available supply capacity, therefore the existing and committed capacity does not need to be subtracted from the maximum constraint, but rather added to ensure the weight vector per region includes the minimum power constraints applicable to that specific region. This process is defined mathematically below.

The constraints for region 1 can be defined mathematically as

$$\boldsymbol{X}_1^w = \{x_{mr}^w \mid m = 2,3,6,7,9,10 \text{ and } r = 1\}, \tag{4.45}$$

$$\boldsymbol{X}_1^s = \{x_{mr}^s \mid m = 3,4,5,8 \text{ and } r = 1\} \tag{4.46}$$

and

$$P_c^w X_1^w + P_c^s X_1^s \leq 1.1\ GW + P_{1\,min}. \tag{4.47}$$

The constraints for region 2 can be defined mathematically as

$$\boldsymbol{X}_2^w = \{x_{mr}^w \mid m = 2,3,6,7,10 \text{ and } r = 2\}, \tag{4.48}$$

$$\boldsymbol{X}_2^s = \{x_{mr}^s \mid m = 3,4,5 \text{ and } r = 2\} \tag{4.49}$$

and

$$P_c^w X_2^w + P_c^s X_2^s \leq 1.74\ GW + P_{2\,min}. \tag{4.50}$$

The constraints for region 3 can be defined mathematically as

$$\boldsymbol{X}_3^w = \{x_{mr}^w \mid m = 1,4,8 \text{ and } r = 3\}, \tag{4.51}$$

$$\boldsymbol{X}_3^s = \{x_{mr}^s \mid m = 2,6,10 \text{ and } r = 3\}, \tag{4.52}$$

and

$$P_c^w X_3^w + P_c^s X_3^s \leq 1.26\ GW + P_{3\,min}. \tag{4.53}$$

The constraints for region 4 can be defined mathematically as

$$\boldsymbol{X}_4^w = \{x_{mr}^w \mid m = 1,4,8 \text{ and } r = 4\}, \tag{4.54}$$

$$\boldsymbol{X}_4^s = \{x_{mr}^s \mid m = 2,6 \text{ and } r = 4\} \tag{4.55}$$

and

$$P_c^w X_4^w + P_c^s X_4^s \leq 4.051\ GW + P_{4\,min}. \tag{4.56}$$

The constraints for region 5 can be defined mathematically as

$$\boldsymbol{X}_5^w = \{x_{mr}^w \mid m = 1,2,3,4,5,6,7,8,10 \text{ and } r = 5\}, \quad 4.57$$

$$\boldsymbol{X}_5^s = \{x_{mr}^s \mid m = 1,2,3,7,9 \text{ and } r = 5\} \quad 4.58$$

and

$$P_c^w X_5^w + P_c^s X_5^s = 0\ GW + P_{5\,min}. \quad 4.59$$

The total number of wind subclusters within each region summates to 29 and the number of solar sub-clusters summates to 18. The research space is therefore increased from 20 to 47 clusters. After the capacity allocation weights have been assigned to each cluster, the results are accumulated to achieve a single weighting for each cluster, the minimum constraints per region are also included in weight vector.

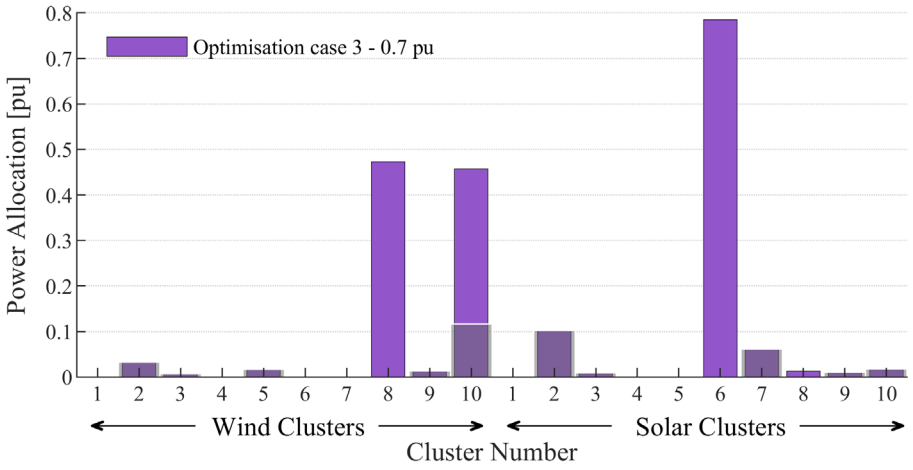
Optimisation case study 3 is implemented to showcase the framework capability, where an increase in computing complexity is seen due to the nature of the maximum constraints.

The constrained optimisation case study 3 determines the spatial capacity allocation scenario for optimal power balance and grid support while considering minimum and maximum capacity constraints defined by the existing wind and solar generation units within the clustered regions. The optimisation framework is implemented using the average ToU clusters as input. The resultant residual load histogram is targeted, where unweighted occurrences above a set threshold of 0.7 pu are minimised.

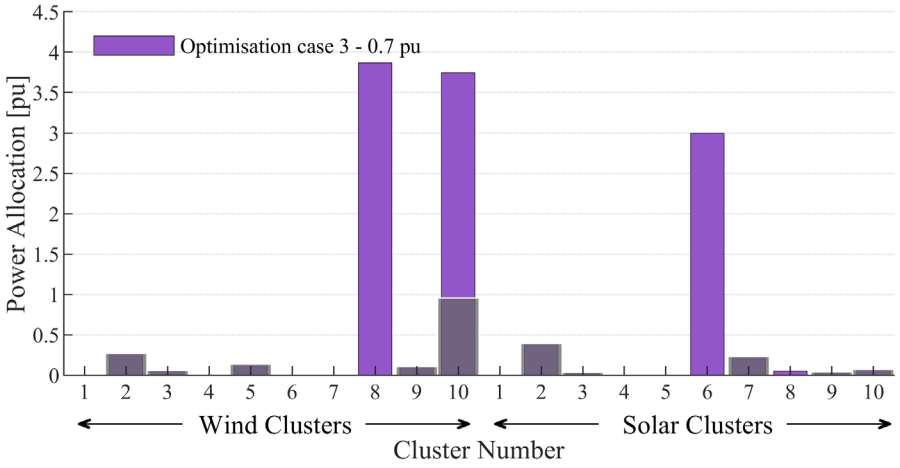
Figure 4.19(a) shows the normalised geospatial capacity allocations and Figure 4.19(b) shows the spatial capacity allocations translated to the power allocations in GW, for the threshold above 0.7 pu power. The capacity allocation for the wind resource is assigned predominately to clusters 8 and 10. The allocation shows that the maximum available capacity, which falls within regions 1 and 2 is allocated to cluster 10. Regions 1 and 2 have a combined maximum constraint 2.84 GW and the combined minimum constraints for cluster 10 is 1.0218 GW, as indicted by the grey bars in Figure 4.19(b). The rest of the capacity is allocated to cluster 8, where the optimisation specifically allocated the capacity to the unconstraint region 6. This allocation will have been the next best cluster in terms of the wind resource characteristics. Looking at Figure 3.14(a)-(c), the wind resource profiles depicted per cluster show that cluster 8 is indeed the best yielding cluster when compared to the other possible options in the unconstraint region 6, namely, clusters 1 and 4.

All unconstraint capacity allocation for the solar resource is assigned to cluster 6. Figure 3.24 shows that cluster 6 is situated in Emalahleni, which is the REDZs within the unconstraint region 6. Figure 3.27 shows that cluster 6 has the highest yield during the morning peak period in the high demand season and it is the best performing cluster when compared with the other cluster allocation option, namely cluster 2.

Figure 4.20 depicts the histogram of the normalised residual load profile obtained with optimisation case study 3, for a set threshold of 0.7 pu power. The histogram of the original demand profile is shown to provide a basis for visual comparison. Optimisation case 3 depicts a maximum frequency of occurance at approximately 0.71 pu power.



(a) Normalised geospatial capacity allocations



(b) Geospatial capacity allocations in GW

Figure 4.19: The geospatial (a) normalised and (b) power capacity allocations for optimisation case study 3, for a set threshold of 0.7 pu power.

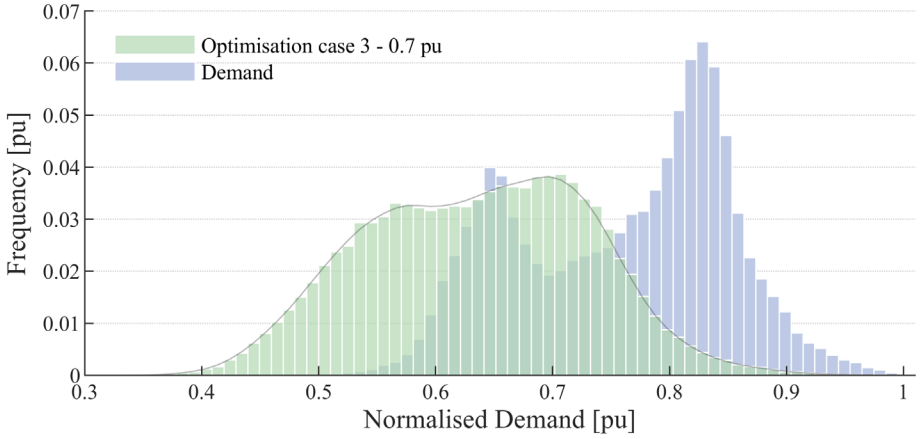
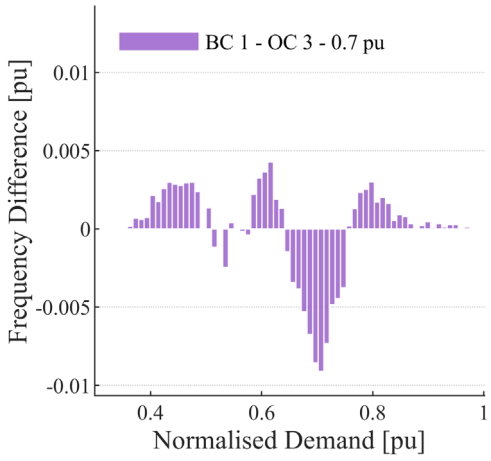


Figure 4.20: Histogram of the optimised residual load profiles for optimisation case studies 3, unweighted objective function with a threshold value of 0.7 pu.

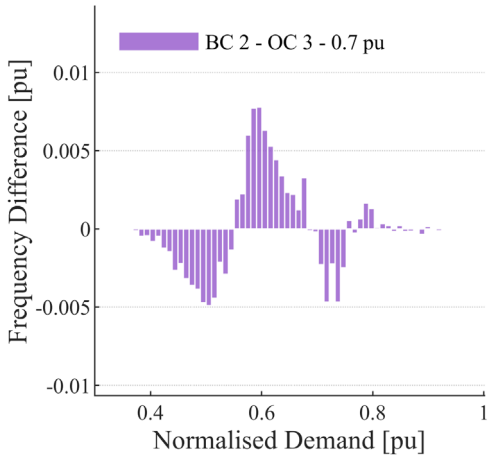
To evaluate the performance of the optimised capacity allocation, the residual load histogram for the optimisation case study 3 is subtracted from each of the baseline case study histograms. This yields bar charts that display the differences in frequency of occurrences between the optimisation case study results and the baseline case studies, above the set threshold of 0.7 pu. However, the minimum and maximum constraint scenario cannot be fairly compared with the baseline case studies since there are major limitations enforced on the capacity allocations.

In Figure 4.21(a)-(c) all of the instances still display positive differences in the higher power ranges, mainly above approximately 0.8 pu. Figure 4.21(c) shows the most improvement, starting from approximately 0.55 pu. Figure 4.21(a)-(d) shows less of an improvement in the frequency of occurrence for the higher residual load power values, which is due to the major maximum constraints applied to the various regions.

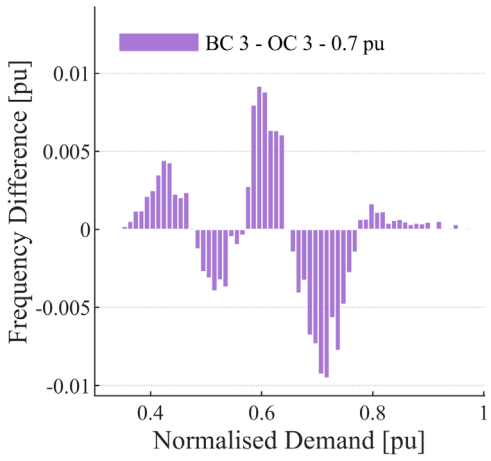
Overall, the results indicate that the optimisation framework is still able to successfully reduce the frequency of occurrence of residual load power values above the set threshold when compared to all baseline case studies, despite the stringent maximum constraints applied.



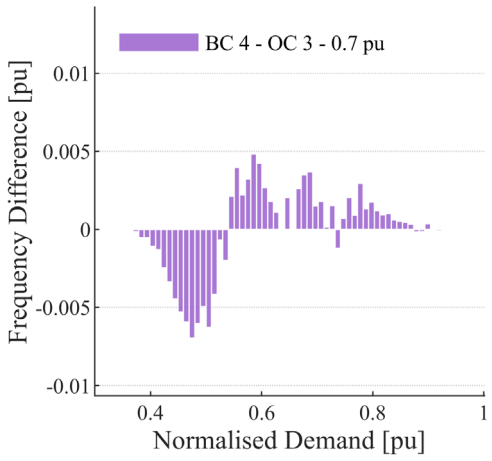
(a) Baseline case study 1: Maximise energy yield for assignment to REDZs profiles



(b) Baseline case study 2: Equal assignment to averaged REDZs profiles



(c) Baseline case study 3: Maximise energy yield for assignment to temporal clusters



(d) Baseline case study 4: Equal assignment to ToU cluster profiles

Figure 4.21: Difference in the histograms of unweighted optimisation case study 3 and the baseline case studies.

Table 4.10 provides the performance evaluation results for the optimisation case study 3. The results show an increase in the frequency of occurrence of residual load power values above the set threshold when compared to optimisation case studies 1 and 2. This is to be expected as stringent constraints have been applied to the capacity allocations. Due to these constraints, it seems unfair to compare the results with the baseline case studies, however the results are still calculated to see the effects of the maximum constraint allocations. Table 4.11 lists the differences between the baseline case study results and the optimisation case study results. Table 4.11 shows a mix of positive and negative values, which shows that the optimisation case study 3 was still able to outperform some baseline case scenarios. The greatest improvement is again seen between baseline case study 4 and the optimisation case study for a threshold of 0.7 pu power. There is a 1.62 % improvement in the frequency of occurrence of residual power values and a 2.14 % improvement in the cumulative energy yield when compared to baseline case study 4. The difference in the frequency of occurrences of residual power values above 90 % of the maximum residual load value, between the optimisation case 3 and baseline case study 4 shows a reduction of 32 instances.

From these results, it is clear that the grid constraints limit the opportunity for optimisation of the geospatial capacity allocations. This has serious implications for future deployment of RE in SA, since our gird is facing severe supply limitations for all added generation capacity.

Table 4.10: Performance of the optimisation case study 3, unweighted objective function with a threshold value of 0.7 pu.

Optimisation case study	F_1^T [%]	F_2^T [%]	F_3^T
3	27.925	36.866	182

Table 4.11: Difference between performance results between the baseline case studies and the optimisation case study 3 for the unweighted case with a set threshold of 0.7 pu.

Difference	F_1^T [%]	F_2^T [%]	F_3^T
1	-1.002	-1.322	9.000
2	-1.214	-1.603	-10.000
3	-3.019	-3.985	8.000
4	1.620	2.139	32.000

4.5 Conclusion: Spatial capacity allocation framework

A strategy is proposed, whereby spatiotemporal wind and solar power profiles are statistically clustered such that the cluster regions represent temporal profiles with the same statistical characteristics for a set of pre-defined ToU periods. Thereafter, the averaged temporal profiles of the clusters are used as the input for a geospatial capacity allocation optimisation procedure, using a statistical risk-based objective function. The objective function targets the frequency of occurrence of the residual load values above a given threshold.

The proposed framework has been implemented for a number of baseline case studies and optimisation case studies. It is concluded that the framework is highly flexible in the sense that the formulation of the minimum and maximum allocation constraints allow application for real-world scenarios where capacity allocation constraints apply on a regional level.

The results show that the geospatial capacity allocations obtained with optimisation case studies 1 and 2 significantly reduce the frequency of load occurrences when compared with the baseline case studies. These reductions appear modest in terms of the performance metrics. However, the framework targets

the occurrences of the higher residual load values. These occurrences translate nonlinearly to high impacts due to the high capital and operating costs associated with these load values.

Optimisation case 3 is considered a real-world scenario, where stringent constraints are applied. This case study cannot be fairly compared with the baseline case studies. However, the results still show an improvement in the frequency of occurrence of the residual load power values above 0.8 pu. From these results, it is clear that the grid constraints limit the opportunity for optimisation of the geospatial capacity allocations. This has serious implications for RE in SA, since our gird is facing severe supply limitations for all newly added generation capacity.

Overall, the optimisation framework provides a robust method for the geospatial capacity allocation of wind and solar resources. The framework employs a robust way of handling constraint scenarios when considering multiple highly granular resource clusters.

Chapter 5

Conclusions and Recommendations

5.1 Overview

South Africa has displayed a unique energy supply profile over recent years, where the ability to consistently meet the energy demand has been constrained by physical limitations of the current energy supply infrastructure. The inadequate supply infrastructure results in countrywide loadshedding events, where total energy supply within peak ToU periods cannot be met. Low-grade coal, poorly maintained power plants and the impending end of life periods approaching for existing thermal plants adds to the country's energy supply deficit. It is expected that optimised spatial capacity allocation of new build wind and solar plants can assist in addressing the generation capacity constraints in the medium to longer term future.

Figure 5.1 depicts a high-level block diagram of the geospatial capacity allocation framework developed and implemented in this work. The framework incorporates a number of distinct processes, including the translation of temporospatial wind and solar resource profiles to normalised power profiles, the clustering of the temporospatial power profiles into geospatial clusters using a novel ToU statistical feature vector approach, and a constrained spatial allocation optimisation process for wind and solar generation capacities. The remainder of this chapter summarises the conclusions pertaining to the different processes, highlights contributions of an original nature, and presents recommendations for further work. The chapter concludes with a summary of the publication outcomes generated by this research.

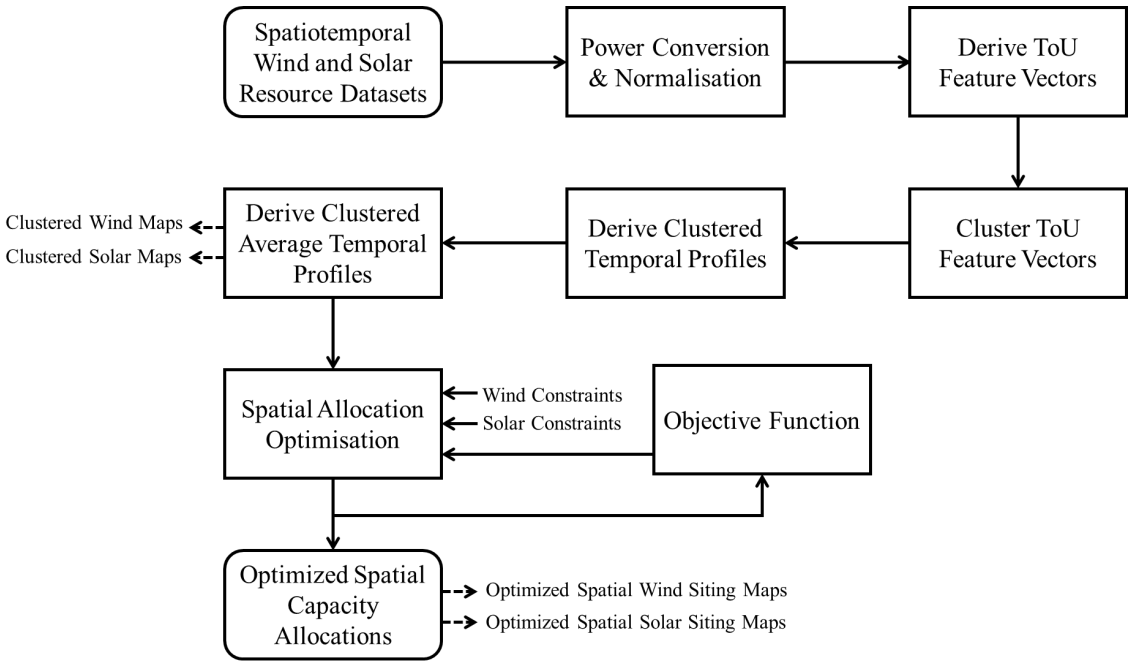


Figure 5.1: Block diagram representation of the geospatial capacity allocation framework proposed in this investigation.

5.2 Research summary and conclusions

5.2.1 Defining the research

Chapter 1 provides a detailed overview of the research background and motivation. The research aim, scope, objectives, and the expected novel contributions are also defined. The main research question is defined as follows:

> *Can a comprehensive RE geospatial capacity allocation framework be developed that incorporates a risk-based approach to grid support, taking cognisance of real-world regional siting constraints, using geospatial wind and solar resource clusters as the input dataset?*

5.2.2 Resource classification model

Clustering represents an integral part of the proposed framework. Clustering of the temporospatial resource data sets into a set of geospatial clusters represents a data reduction methodology, which reduces the computational expense of the optimised spatial capacity allocation process considerably. In practice, clustered resource maps are also more appropriate for plant siting studies compared to resource maps defined in terms of single coordinates, as is predominantly the case for the historical studies reported in literature. This is due to single sites having high risks of choosing coordinates which are not viable for plant placement, or large quantities of RE resources are allocated to coordinates with limited integration capacities.

The cluster distributions obtained with the ToU feature vectors are compared with distributions delivered by clustering the spatiotemporal profiles using the k-means algorithm. This delivers a significant result in the sense that the distributions obtained with the ToU feature vectors across the eleven geographical REDZs, for the same number of clusters, show significantly higher granularity and diversity. It is concluded that the clustering of ToU feature vectors, compared to the clustering of raw temporal profiles, distinguishes more finely between the daily and seasonal temporal properties of neighbouring geographic locations. A statistical analysis of the clustering results obtained with the two clustering approaches are presented. The analysis show that the power profiles associated with the individual clusters obtained using the feature vector approach have similar or higher means, and similar or lower standard deviations, compared to the clusters obtained using spatiotemporal profiles.

The ToU feature vector-based approach decreases the dimensionality of the input resource dataset significantly. Compared to clustering of the temporal profiles, the statistical feature vector approach delivers improved correlation between the geospatial clusters and the underlying topographic and climatic features.

Overall, the ToU based clustering approach proposed and implemented in the study shows excellent performance as a data reduction pre-processing step in optimising the site locations of future wind and solar energy plants for optimal grid support in the context of the diurnal and seasonal characteristics of the aggregated load profile. The ToU periods used in the case study can be readily amended to represent other temporal criteria of interest, such as time-dependent transmission constraints, real-time pricing considerations, the pumping schedules of hydro-storage systems, etc.

5.2.3 Optimal spatial capacity allocation framework

A comprehensive systematic literature review is conducted on the strategies proposed in literature for the optimal spatial capacity allocation of wind and solar resources in the context of varying optimisation

objectives. These contributions are analysed with the view to determine the similarities and differences with the framework proposed in this research.

The integration framework is developed using the best-performing clustering algorithm with the ToU feature-based vectors as input. Thereafter, clustered wind and solar power geographic resource maps are derived for the South African REDZs.

The proposed geospatial capacity allocation framework is formulated mathematically and is implemented for a range of risk-based objective functions defined in terms of the residual load profile. These objective functions include minimising the unweighted frequency of occurrence above a set threshold and minimising the weighted frequency of occurrence above a set threshold. Case study results are obtained for various case study scenarios, including scenarios without capacity allocation constraints, and scenarios with minimum and/or maximum capacity constraints.

The capacity allocations obtained with the proposed framework are compared with the capacity allocations associated with various baseline case studies. These baseline case studies are defined to mimic best case scenarios, as well as various siting strategies that do not include optimisation of the spatial capacity allocations, such as represented by the current REIPPP allocations. The performance evaluation showed a significant decrease in the frequency of occurrence for load values above the set thresholds, when compared to the baseline case study scenarios. The results are detailed as follows:

For all geospatial capacity allocation optimisation case studies, the resultant capacity outputs showed the most improvement when compared with the baseline case study 4. This is a significant result, since baseline case study 4 represents the unoptimised, equal weight assignment scenario, using the same ToU feature-based input profiles as the optimisation case studies.

- *Optimisation case study 1*

  This case study shows a diverse set of optimised capacity allocation weights, with wind cluster 10 and solar clusters 1 and 9, showing large capacity allocation percentages.

  Various evaluation metrics are defined to measure the optimised allocation performances. For this optimisation scenario, the greatest improvement is seen between baseline case study 4 and the unweighted optimisation case study, for a threshold of 0.7 pu power. There is a 7.18 % improvement in the difference between the frequency of occurrence of residual power values and a 9.48 % improvement in the cumulative energy yield when compared to the baseline case 4 scenario. The frequency of occurrences of residual power values above 90 % of the maximum residual load value is reduced by 78 instances.

- *Optimisation case study 2*

  This case study shows similar results to optimisation case study 2 since the only difference in the capacity allocation scenarios is the introduction of minimum capacity constraints. However, solar clusters 9 and 10 now show the largest capacity allocation percentages.

  Results indicate that optimising the geospatial capacity allocation, rather than allocating equally between clusters, successfully aids in reducing the frequency of occurrence of residual load power values above the set thresholds.

  The greatest improvement is again seen between baseline case study 4 and the unweighted optimisation case study for a threshold of 0.7 pu power. There is a 7.05 % improvement in the difference between the frequency of occurrence of residual power values and a 9.3 % improvement in the cumulative energy yield when compared to baseline case study 4. The difference in the frequency of occurrences of residual power values above 90 % of the maximum residual load value, between the optimisation case 2 and base case study 4 shows a reduction of 84 instances. This shows

a significant result since these high residual load values quantify the most expensive operating time when replaced by ancillary services.

– *Optimisation case study 3*

  This case study is implemented to showcase the framework capabilities, where an increase in computing complexity is seen due to the nature of the maximum constraints.

  The constrained optimisation case study 3 determines the spatial capacity allocation scenario for optimal power balance and grid support while considering stringent minimum and maximum capacity constraints.

  The greatest improvement is again seen between baseline case study 4 and the optimisation case study for a threshold of 0.7 pu power. There is a 1.62 % improvement in the difference between the frequency of occurrence of residual power values and a 2.14 % improvement in the cumulative energy yield when compared to baseline case study 4. The difference in the frequency of occurrences of residual power values above 90 % of the maximum residual load value, between the optimisation case 3 and base case study 4 shows a reduction of 32 instances.

  From these results, it is clear that the grid constraints limit the opportunity for optimisation of the geospatial capacity allocations. This has serious implications for RE in SA, since our gird is facing severe supply limitations for all newly added generation capacity.

Overall, in support of a reliable future energy supply scenario with high penetration of RE, the optimisation framework proposed in this work represents a probabilistic risk-based approach that seeks to minimise the number of events where high residual load values require ancillary service interventions to maintain power balance.

5.3 Research contributions

The novel contributions pertaining to the study can be summarised as follows:

– *Time-of-Use feature based clustering methodology:*

  A detailed comparison is conducted of various clustering methodologies for the use on wind and solar resource data, for both statistical and temporal inputs. The RE resource is initially clustered using generic temporal input data sets, which yields clusters formations which align closely within the REDZs, showing low cluster granularity within each REDZs. The temporal input data sets each span a 5-year period, with an hourly resolution for the solar resource and a 15-minute resolution for the wind resource. This summates to 43 824 solar instances for each of the 4470 spatial locations and 175 296 wind instances for each of the 4470 spatial locations. The temporal data sets are translated to statistical feature vectors, which model the high and low demand ToU profiles, which can be adapted to incorporate different criteria, such as time-dependent transmission constraints, real-time pricing considerations, etc. The ToU feature vectors statistically model the daily peak, standard and off-peak periods, which places equal importance on all periods. The spatiotemporal clustering method inherently places more emphasis on the off-peak periods since these periods occur for longer time intervals. The ToU feature-based clustering approach summates the large temporal dataset to two daily representations of the high and low demand seasons. This summates to 12 instances for each of the 4470 spatial locations for the wind resource clustering. The solar resource clusters are further segregated into the morning, midday and evening periods, therefore summating to 26 instances for each of the 4470 spatial locations. The ToU cluster formations depict highly granular cluster formations, which show clear ties to the underlying terrain and climatic conditions.

Overall, the novel ToU feature-based clustering method drastically reduces the dimensionality of the data sets and creates well defined clusters which better capture the underlying characteristics of the RE resource profiles. The clusters show improved granularity and accuracy when compared to clustering methodologies based on temporal profiles.

– *Probabilistic risk-reduced approach:*

The formulated objective functions target the higher load values in the residual load profile histogram in a statistical manner. As such, this method represents a probabilistic risk-based approach that seeks to minimise the number of events where high residual load values require ancillary service interventions to maintain power balance. From a financial perspective, the proposed approach also reduces the capital expenditure requirements associated with baseload generation capacity and ancillary service, as well as the energy costs associated with expensive peaking generation, such as open-cycle gas turbines.

In practice, the cost function can be easily amended to target the probability density function of the residual load profile for a given ToU period, etc. The function can also be readily adapted to implement a non-linear weighting, such as an exponentially increasing weighting function, which would be more effective in targeting extreme values.

– *Robust constraint capabilities:*

In practice, the optimal siting of RE generation is not only defined in terms of power balance, but is subject to considerations such as the availability of grid infrastructure, power evacuation capacity constraints, economic and socio-economic considerations, land use and availability, environmental impacts, etc.

The concept of regional constraints is introduced in order to enforce constraints on the same cluster profiles which fall into different constrained regions. The maximum constraints are often locationally dependent, as they are subject to the capacity limit imposed by nearby substations. In practice, the constraints may translate to a combinational permutation of minimum and/or maximum constraints formulated individually for the wind and solar allocations at cluster level and minimum and/or maximum constraints formulated for the combined wind and solar allocations at cluster level and/or regional level.

The proposed constraint approach can handle a complex set of capacity allocation constraints in a robust manner. However, constraints that are defined at a regional level require that the optimisation algorithm take cognisance of the individual subclusters present in those regions. This increases the dimensions of the search space, which gives rise to longer simulation times and increased challenges with problems related to phenomena such as local minima.

– *Optimised spatial capacity allocation framework*:

The research reported in literature, with a few exceptions that are only applicable for specific case study scenarios, confirms that geospatial optimisation of RE capacity allocation, especially in the context of the spatiotemporal properties of the RE resources, can contribute in supporting power balance and various other operational objectives at high levels of RE penetration. Overall, it is concluded that, compared to the body of work reported in literature, the proposed RE geospatial capacity allocation framework proposed in this dissertation exhibits aspects of novelty and originality. This is particularly in terms of the pre-classification of the RE resource data, the risk-based objective function, and the use of adaptable optimisation constraints to accommodate real-world scenarios.

5.4 Recommendations for future work

This research has uncovered potential opportunities for further investigation. These can be summarised as follows:

- Increase the number of clusters to see the effects on the spatial capacity allocations.
- Increase the study region to include the whole of South Africa, including all of the existing RE generation capacities.
- Implement a non-linear weighting function in the objective function, such as an exponentially increasing the weighting function, which would be more effective in targeting extreme values.
- Implement different optimisation methods with the view to decrease the simulation times reduce the risks associated with local minima for highly dimensional input datasets.
- Consider hydrogen plants to alleviate the stringent maximum regional allocation constraints currently imposed by the power evacuation limitations. In this approach, excess capacity is diverted into hydrogen plants instead of curtailing excess RE generation.

5.5 List of research publications

This research has generated the following publications.

5.5.1 Conference articles

- Clustered wind resource domains for the South African Renewable Energy Development Zones, C. Y. Janse van Vuuren, H. J. Vermeulen, 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA).
- Clustering of wind resource Weibull characteristics on the South African renewable energy development zones, C. Y. Janse van Vuuren, H. J. Vermeulen, J. C. Bekker, 2019 10th International Renewable Energy Congress (IREC).
- Optimised Geographical Allocation of Wind Energy Capacity using a Mean-Variance Portfolio Algorithm for Clustered and Un-clustered Profiles, C. Y. Janse van Vuuren, H. J. Vermeulen, J. C. Bekker, 2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe).
- Optimal Siting of Wind Energy Capacity for Minimum Residual Load Variance, C. Y. Janse van Vuuren, H. J. Vermeulen, M. Groch, 2019 9th International Conference on Power and Energy Systems (ICPES). (Best paper award)
- Optimal Geospatial Allocation of Wind Generation Capacity for Grid-Support Objectives, C. Y. Janse Van Vuuren, H. J. Vermeulen, M. Groch, 2020 11th International Renewable Energy Congress (IREC), 1-6.
- A Self-Organizing Map Approach for Time-of-Use Feature Based Wind Resource Clustering, C.Y. Janse Van Vuuren, H. J. Vermeulen, M. Groch, 2020 11th International Renewable Energy Congress (IREC). (Best paper award)

5.5.2 Journals articles

- Clustering of wind resource data for the South African renewable energy development zones, C. Y. Janse van Vuuren, H. J. Vermeulen, Journal of Energy in Southern Africa 30(2), 126-143.
- Wind resource clustering based on statistical Weibull characteristics, C. Y. Janse van Vuuren, H. J. Vermeulen, Wind Engineering 43(4), 359-376.
- A statistical Time-Of-Use tariff based wind resource clustering approach using Self-Organizing Maps, C. Y. Janse van Vuuren, H. J. Vermeulen, M Groch, Wind Engineering 45(4), 807-821.
- Time of Use Feature Based Clustering of Spatiotemporal Wind Power Profiles, C. Y. van Staden, H. J. Vermeulen, M. Groch, Energy Elsevier 236.

References

[1] Council for Scientific and Industrial Research, "National wind solar sea," 17 February 2016. [Online]. Available: https://www.csir.co.za/national-wind-solar-sea. [Accessed 29 July 2019].

[2] California Independent System Operator Corporation, "Impacts of renewable energy on grid operations," California ISO, California , 2017.

[3] P. Denholm, M. O'Connell, G. Brinkman and J. Jorgenson , "Overgeneration from Solar Energy in California: A Field Guide to the Duck Chart," National Renewable Energy Laboratory, California, 2015.

[4] P. del Río and L. Janeiro, "Overcapacity as a Barrier to Renewable Energy Deployment: The Spanish Case," *Journal of Energy,* pp. 1-10, 2016.

[5] A. Couto and A. Estanqueiro, "Exploring wind and solar PV generation complementarity to meet electricity demand," *Energies,* vol. 13, no. 16, 2020.

[6] C. E. Hoicka and I. H. Rowlands, "Solar and wind resource complementarity: Advancing options for renewableelectricity integration in Ontario, Canada," *Renewable Energy,* vol. 36, pp. 97-107, 2011.

[7] E. Ela, M. Milligan and B. Kirby, "Operating Reserves and Variable Generation," National Renewable Energy Laboratory, Colorado, 2011.

[8] M. Hedayati-Mehdiabadi, J. Zhang and K. W. Hedman, "Wind Power Dispatch Margin for Flexible Energy and Reserve Scheduling With Increased Wind Generation," *IEEE Transactions on Sustainable Energy,* vol. 6, no. 4, pp. 1543-1552, 2015.

[9] J. R. Calitz and J. G. Wright, "Statistics of utility-scale power generation in South Africa in 2020," in *The CSIR Energy Centre on the statistics of utility-scale power generation in South Africa in 2020*, 2021.

[10] The Danish Energy Agency, "Denmark's Energy and Climate Outlook," The Danish Energy Agency, Denmark, 2018.

[11] Union of Concerned Scientists , "Turning Down the Gas in California," Union of Concerned Scientists , Cambridge, 2018.

[12] California Energy Commission, California Public Utilities Commission, California Air Resources Board, "SB 100 Joint Agency Report: Charting a path to a 100% Clean Energy Future," California Energy Commission, California , 2021.

[13] Department of Energy: South Africa, "Integrated Resource Plan Draft," Department of Energy, South Africa, Pretoria, 2018.

[14] Department of Energy: South Africa, "Integrated Resource Plan (IRP2019)," Department of Mineral Resources and Energy, Pretoria, 2019.

[15] A. Eberhard and R. Naude, "The South African renewable energy IPP procurement programme," Graduate School of Business, University of Cape Town, Cape Town, 2017.

[16] Mineral resources and energy, “Independent Power Producers Procurement Programme (IPPPP) An Overview,” Independent Power Producers Office, Johannesburg, South Africa, 2019.

[17] Mineral resources and energy, “Independent Power Producers Procurement Programme (IPPPP): An Overview,” Independent Power Producer Office, Johannesburg, South Africa, 2021.

[18] A. Eberhard and R. Naude, “The South African Renewable Energy IPP Procurement Programme: Review, Lessons Learned & Proposals to Reduce Transaction Costs,” Graduate School of Business, University of Cape Town, Cape Town, 2017.

[19] Department of Environment, Forestry and Fisheries, “Government Gazette No. 44191 Vol. 668,” 26 February 2021. [Online]. Available: sapvia.co.za/wp-content/uploads/2021/03/44191_26-02_NationalGovernment-REDACT2.pdf. [Accessed 2021 05 2021].

[20] M. Liserre, T. Sauter and J. Y. Hung, “Future Energy Systems,” in *IEEE Industrial Electronics Magazine*, 2010.

[21] APS Panel on Public Affairs, “Integrating Renewable Electricity on the Grid,” American Physical Society, Washington DC, 2007.

[22] Greening the Grid, “Using Wind and Solar to Reliably Meet Electricity Demand,” U.S. Government’s Enhancing Capacity for Low Emission Development Strategies (EC-LEDS) program, 2015.

[23] J. Hossain and H. R. Pota, Robust Control for Grid Voltage Stability: High Penetration of Renewable Energy, Springer, 2014.

[24] Q. Hou, N. Zhanga, E. Du, M. Miao, F. Peng and C. Kanga, “Probabilistic duck curve in high PV penetration power system: Concept, modeling, and empirical analysis in China,” *Applied Energy, Elsevier,* vol. 242, pp. 205-215, 2019.

[25] C. W. Potter, A. Archambault and K. Westrick, “Building a Smarter Grid through Better Renewable Energy Information,” in *IEEE/PES Power Systems Conference and Exposition*, Seattle, USA, 2009.

[26] G. M. Shafiullah, M. T. Amanullah, D. Jarvis, S. Ali and P. Wolfs, “Potential Challenges: Integrating Renewable Energy with the Smart Grid,” in *20th Australasian Universities Power Engineering Conference*, Christchurch, New Zealand, 2010.

[27] H. H. Ei-Tamaly and M. A. Wahab, “Simulation of Directly Grid-Connected Wind Turbines for Voltage Fluctuation Evaluation,” *Journal of Applied Engineering Research,,* vol. 2, no. 1, pp. 15-30, 2007.

[28] E. Bossanyi, Z. Saad-Saoud and N. Jenkins, “Prediction of Flicker Produced by Wind Turbines,” *Wind Energy,* vol. 1, pp. 35-51, 1998.

[29] N. T. Linh, “Power Qulaity Investigation of Grid Connected Wind Turbines,” in *4th IEEE Conference on Industrial Elctronics and Applications,*, China, 2009.

[30] Eaton Electrical , “Harmonics in your electrical system,” Eaton Corporation .

[31] Z. Ming, H. Lixin, Y. Fam and J. Danwei, "Research of the problems of renewable energy orderly combined to the grid in smart grid," in *Power and Energy Engineering Conference (APPEEC 2010)*, Chengdu, 2010.

[32] R. Billinton and R. N. Allan, "Power-system reliability in perspective," *Electronics & Power* , pp. 231-236, 1984.

[33] R. Billinton and D. Huang, "Basic considerations in generating capacity adequacy evaluation," in *Canadian Conference on Electrical and Computer Engineering*, Saskatoon, Canada, 2005.

[34] E. Ibanez and M. Milligan, "Comparing Resource Adequacy Metrics," in *13th International Workshop on Large-Scale Integration of Wind Power into Power Systems as Well as on Transmission Networks for Offshore Wind Power Plants*, Berlin, 2014.

[35] L. Söder and M. Amelin, "A review of different methodologies used for calculation of wind power capacity credit," in *Proc. IEEE Power Eng. Soc. General Meeting*, Pittsburgh, 2008.

[36] M. Amelin, "Comparison of Capacity Credit Calculation Methods for Conventional Power Plants and Wind Power," *IEEE Transactions on Power Systems,* vol. 24, no. 2, pp. 685-691, 2009.

[37] Eskom, "Tariffs & Charges 2019/2020," Eskom Holdings SOC Ltd, Pretoria, South Africa, 2020.

[38] G. C. Wu, R. Deshmukh, K. Ndhlukula, T. Radojicic, J. Reilly-Moman, A. Phadke, D. M. Kammen and D. S. Callaway, "Strategic siting and regional grid interconnections key to low-carbon futures in African countries," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 114, no. 15, p. 3004–3012, 2017.

[39] The International Renewable Energy Agency, "Planning for the Renewable future," The International Renewable Energy Agency, Abu Dhabi, 2017.

[40] T. Aboumahboub, K. Schaber, P. Tzscheutschler and T. Hamacher, "Optimization of the Utilization of Renewable Energy Sources in the Electricity Sector," in *Recent Advances in the Analysis of Sustainable Energy and Environment*, 2010.

[41] S. Beckera, B. A. Frew, G. B. Andresen, T. Zeyer, A. Schramm, M. Greiner and M. Z. Jacobson, "Features of a fully renewable US electricity system: Optimized mixes of wind and solar PV and transmission grid extensions," *Elsevier,* 2014.

[42] H. Takeda, P. Veerkamp, T. Tomiyama and H. Yoshikawa, "Modeling Design Processes," *AI Magazine,* vol. 11, no. 4, 1990.

[43] CSIR, CSIR Report Number: CSIR/SPLA/SECO/ER/2019/0085, Phase 1 Strategic Environmental Assessment for wind and solar PV energy in South Africa. CSIR Report Number: CSIR/SPLA/SECO/ER/2019/0085 Stellenbosch, Western Cape., Stellenbosch, Western Cape: Department of Environment Forestry and Fisheries, 2015.

[44] Department of Environmental Affairs, "Strategic Environmental Assessment for wind and solar photovoltaic energy in South Africa," CSIR Report Number: CSIR/CAS/EMS/ER/2015/0001/B, Stellenbosch, 2015.

[45] J. Badger, H. Frank, A. N. Hahmann and G. Giebel, "Wind-Climate Estimation Based on Mesoscale and Microscale Modeling: Statistical-Dynamical Downscaling for Wind Energy Applications," *Journal of Applied Meteorology and Climatology,* vol. 53, p. 1901–1919, 2014.

[46] Fraunhofer IWES and The CSIR Energy Centre, "Wind and Solar PV Resource Aggregation Study for South Africa," Fraunhofer IWES, South Africa, 2016.

[47] Department of Mineral Resources and Energy South Africa, Sonedi, Royal Danish Embassy and Wind Atlas for South Africa, "The WASA download site," Council for Scientific and Industrial Research, Pretoria, 2010.

[48] R. Müller, U. Pfeifroth, C. Tr€ager-Chatterjee, R. Cremer, J. Trentmann and R. Hollmann, "Surface Solar Radiation Data Set - Heliosat (SARAH) - Edition 1," *Satellite Application Facility on Climate Monitoring, CM SAF,* 2015.

[49] G. M. Masters, "Wind Farms," in *Renewable and Efficient Electric Power Systems*, New Jersey, JOHN WILEY & SONS, INC., 2004, pp. 351-360.

[50] Vestas, "Vestas 3 MW Platform," 15 10 2015. [Online]. Available: https://www.nhsec.nh.gov/projects/2013-02/documents/131212appendix_15.pdf. [Accessed 24 06 2021].

[51] H. Nørgaard and H. Holttinen, "A Multi-Turbine Power Curve Approach," in *Nordic Wind Power Conference*, Gothenburg, 2004.

[52] T. Huld, G. Friesen, A. Skoczek, R. P. Kenny, T. Sample, M. Field and E. D. Dunlop, "A power-rating model for crystalline silicon PV modules," *Solar Energy Materials & Solar Cells,* vol. 95, pp. 3359-3369, 2011.

[53] D. L. King, W. E. Boyson and J. A. Kratochvil, "Photovoltaic Array Performance," Sandia National Laboratories, 2004.

[54] D. Faiman, "Assessing the outdoor operating temperature of photovoltaic modules," *Progress in Photovoltaics: Research and Applications ,* vol. 16, p. 307–315, 2008.

[55] M. Koehl, M. Heck, S. Wiesmeier and J. Wirth, "Modeling of the nominal operating cell temperature based on outdoor weathering," *Solar Energy Materials & Solar Cells,* vol. 95, p. 1638–1646, 2011.

[56] A. Zagouras, R. H. Inman and C. F. M. Coimbra, "On the determination of coherent solar microclimates for utility planning and operations," *Solar Energy,* vol. 102, pp. 173-188, 2014.

[57] F. Vallee, G. Brunieau, M. Pirlot, O. Deblecker and J. Lobroy, "Optimal wind clustering methodology for adequacy evaluation in system generation studies using nonsequential Monte Carlo simulation," *IEEE Transactions on Power Systems,* vol. 26, no. 4, pp. 2173-2184, 2011.

[58] M. Burlando, M. Antonelli and C. F. Ratto, "Mesoscale wind climate analysis: identification of anemological regions and wind regimes," *Int. J. Climatol.,* vol. 28, no. 5, pp. 629-641, 2008.

[59] F. Cassola, M. Burlando, M. Antonelli and C. F. Ratto, "Optimization of the Regional Spatial Distribution of Wind Power Plants to Minimize the Variability of Wind Energy Input into Power Supply Systems," *American Meteorology Society,* vol. 47, pp. 3099-3116, 2008.

[60] L. L. Tupper, D. S. Matteson and C. L. Anderson, "Band Depth Clustering for Nonstationary Time Series and Wind Speed Behavior," in *Technometrics*, 2015.

[61] T. Leenman and F. Phillipson, "Optimal Placing of Wind Turbines: Modelling the Uncertainty," *J. Clean Energy Technol.,* vol. 3, no. 2, pp. 91-105, 2015.

[62] R. Snel and H. Lundstedt, "Self organising maps of solar wind structures," in *Proceedings of Artificial Intelligence Applications*, Sweden, 1993.

[63] S. Berkovic, "Wind regimes and their relation to synoptic variables using self-organizing maps," *Advances in Science and Research,* vol. 15, pp. 1-9, 2018.

[64] M. Yesilbudak, "Clustering analysis of multidimensional wind speed data using k-means approach," in *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, Birmingham, 2016.

[65] J. G. Pinto, C. P. Neuhaus, G. C. Leckebusch, M. Reyers and M. Kerschgens, "Estimation of wind storm impacts over Western Germany under future climate conditions using a statistical—dynamical downscaling approach," *Dynamic Meteorology and Oceanography,* vol. 62, no. 2, pp. 188-201, 2010.

[66] L. Dong, "Wind Resource Assessment in the Southern Plains of the US: Characterizing Large-Scale Atmospheric Circulation with Cluster Analysis," *Atmosphere,* vol. 9, no. 3, p. 2018, 2018.

[67] P. Nahmmacher, E. Schmid, L. Hirth and B. Knopf, "Carpe diem: A novel approach to select representative days for longterm power system modeling," *Energy,* vol. 112, pp. 430 - 442, 2016.

[68] B. Andreopoulos, A. An and X. Wang, "A roadmap of clustering algorithms: finding a match for a biomedical application," *Brief. Bioinform.,* vol. 10, no. 3, pp. 297-314, 2009.

[69] A. Bhat, "k-medoids clustering using partitioning around medoids for performing face recognition," *International Journal of Soft Computing, Mathematics and Control (IJSCMC),* vol. Vol. 3, no. No.3, 2014.

[70] S. Ayramo and T. Karkkainen, "Introduction to partitioning-based clustering methods with a robust example," University of Jyvaskyla Department of Mathematical Information Technology, Jyvaskyla, 2006.

[71] Z. Guo-fu and Q. U. Gou-qing, "Analysis and implementation of CLARA algorithm on clustering," *Journal of Shandong University of Technology (Science and Technology),* vol. 02, 2006.

[72] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function, MA, USA: Kluwer Academic Publishers Norwell, 1981.

[73] D. Dembélé and P. Kastner, "Fuzzy C-means method for clustering microarray data," *Bioinformatics ,* vol. 19, no. 8, p. 973–980, 2003.

[74] M. Ester, H. P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Kdd,* vol. 96, no. 34, p. 226–231, 1996.

[75] D. H. Fisher, "Knowledge acquisition via incremental conceptual clustering," *Mach. Learn,* vol. 2, no. 2, pp. 139-172, 1987.

[76] T. W. Liao, "Clustering of time series data—a survey," *Pattern Recognition,* vol. 38, no. 11, pp. 1857-1874, 2005.

[77] S. Rani and G. Sikka, "Recent techniques of clustering of time series data: a survey," *International Journal of Computer Applications ,* vol. 52, no. 15, pp. 1-9, 2012.

[78] J. Lin, M. Vlachos, E. Keogh and D. Gunopulos, "Iterative incremental clustering of time series," in *EDBT*, 2004, pp. 521-522.

[79] S. Aghabozorgi, A. S. Shirkhorshidi and W. Y. Teh, "Time-series clustering – A decade review," *Information Systems,* vol. 53, pp. 16-38, 2005.

[80] W. He, G. Feng, Q. Wu, T. He, S. Wan and J. Chou, "A new method for abrupt dynamic change detection of correlated time series," *International Journal of climatology,* vol. 32, no. 10, pp. 1604-1614, 2012.

[81] D. Graves and W. Pedrycz, "Proximity fuzzy clustering and its application to time series clustering and prediction," in *Intelligent Systems Design and Applications ISDA10*, 2010.

[82] N. Pavlidis, V. P. Plagianakos, D. K. Tasoulis and M. N. Vrahatis, "Financial forecasting through unsupervised clustering and neural networks," *Operational Research,* vol. 6, no. 2, p. 103–127, 2006.

[83] A. Sfetsos and C. Siriopoulos, "Time series forecasting with a hybrid clustering scheme and pattern recognition," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* vol. 34, no. 3, p. 399–405, 2004.

[84] T. Mitsa, Temporal Data Mining, vol. 33, New York: Chapman and Hall/CRC, 2010.

[85] E. Ghysels, P. Santa-Clara and R. Valkanov, "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics,* vol. 131, no. 1-2, pp. 59-95, 2006.

[86] E. Keogh, "Hot sax: efficiently finding the most unusual time series," in *Fifth IEEE International Conference on Data Mining ICDM05*, 2005.

[87] G. Duan, Y. Suzuki and K. Kawagoe, "Grid representation of time series data for similarity search," in *The institute of Electronic, Information,and Communication Engineer*, 2006.

[88] E. J. Keogh, M. J. Pazzani, K. Chakrabarti and S. Mehrotra, "A simple dimensionality reduction technique for fast similarity search in large time series databases," *Knowledge Discovery and Data Mining. Current Issues and New Applications ,* vol. 1805, pp. 122-133, 2000.

[89] H. Zhang and V. Lesser, "Multi-agent based peer-to-peer information retrieval systems with concurrent search sessions," in *Autonomous agents and multiagent systems*, Japan, 2006.

[90] J. Lin, E. Keogh, L. Wei and S. Lonardi, "Experiencing SAX: a novel symbolic representation of time series," *Data Mining and Knowledge Discovery,* vol. 15, no. 2, pp. 107-144, 2007.

[91] K. Chan and A. W. Fu, "Efficient time series matching by wavelets," *15th International Conference on Data Engineering,* vol. 15, no. 3, p. 126–133, 1999.

[92] B. K. Yi and C. Faloutsos, "Fast time sequence indexing for arbitrary Lp norms," in *26th International Conference on Very Large Data Bases*, 2000.

[93] Y. L. Wu, D. Agrawal and A. El Abbadi, "Comparison of DFT and DWT based similarity search in time-series databases," in *Ninth International Conference on Information and Knowledge Management*, 2000.

[94] H. Sakoe and S. Chiba, "A dynamic programming approach to continuous speech recognition," *Seventh International Congress on Acoustics,* vol. 3, pp. 65-69, 1971.

[95] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust. Speech Signal Process,* vol. 26, no. 1, pp. 43-49, 1978.

[96] J. L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *Am. Stat,* vol. 42, no. 1, pp. 59-66, 1988.

[97] C. Faloutsos, M. Ranganathan and Y. Manolopoulos, "Fast subsequence matching in time-series databases," *ACM SIGMOD Rec,* vol. 23, no. 2, pp. 419-429, 1994.

[98] A. Kassambara, "Clustering Distance Measures," in *Practical Guide To Cluster Analysis in R*, STHDA, 2017, pp. 25-27.

[99] E. Keogh and S. Kasetty, "On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration," *Data Mining and Knowledge Discovery,* vol. 7, no. 4, pp. 349-371, 2003.

[100] C. Ratanamahatana and E. Keogh, "Three myths about dynamic time warping data mining," in *International Conference on Data Mining*, 2005.

[101] C. Y. Janse van Vuuren and H. J. Vermeulen, "Clustered wind resource domains for the south african renewable energy development zones," Cape Town, 2019.

[102] R. L. Thorndike, "Who Belongs in the Family?," *Psychometrika,* vol. 18, no. 4, p. 267–276, 1953.

[103] C. A. Sugar and G. M. James, "Finding the number of clusters in a data set: An information theoretic approach," *Journal of the American Statistical Association,* vol. 98, p. 750–763, 2003.

[104] A. Bhat, "k-medoids clustering using partitioning around medoids for performing face recognition," *International Journal of Soft Computing, Mathematics and Control,* vol. 3, no. 3, 2014.

[105] C. Martha, W. Milligan and G. Cooper, "Methodology review: clustering methods," *Applied psychological measurement,* vol. 11, no. 4, pp. 329-354, 1987.

[106] L. Scrucca, M. Fop, T. B. Murphy and A. E. Raftery, "mclust 5: clustering, classification and density estimation using gaussian finite mixture models," *NCBI,* vol. 8, no. 1, p. 289–317, 2016.

[107] R. Boyles, "On the convergence of the EM algorithm," *Journal of the Royal Statistical Society,* vol. 45, p. 47–50, 1983.

[108] T. Kohonen, "Adaptive, associative, and self-organizing functions in neural computing," *Applied Optics ,* vol. 26, no. 23, pp. 4910-4918, 1987.

[109] T. Kohonen, "MATLAB Implementations and Applications of the Self-Organizing Map," Unigrafia Oy, Helsinki, Finland, 2014.

[110] S. K. Kingrani, M. Levene and D. Zhang, "Estimating the number of clusters using diversity," *Artificial Intelligence Research,* vol. 7, no. 1, pp. 15-22, 2018.

[111] M. Yan and K. Ye, "Determining the Number of Clusters Using the Weighted Gap Statistic Statistic," *The International Biometric Society,* vol. 63, no. 4, pp. 1031-1037, 2007.

[112] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *International Journal of Advance Research in Computer Science and Management Studies,* vol. 1, no. 6, 2013.

[113] M. A. Syakur, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster," *IOP Conference Series: Materials Science and Engineering,* vol. 336, 2018.

[114] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics,* vol. 20, pp. 53-65, 1987.

[115] B. Kim, J. Kim and G. Yi, "Analysis of clustering evaluation considering features of item response data using data mining technique for setting cut-off scores," *Symmetry MDIP,* vol. 9, no. 5, 2017.

[116] T. Thinsungnoena, N. Kaoungku, P. Durongdumronchai, K. Kerdprasop and N. Kerdprasop, "The Clustering Validity with Silhouette and Sum of Squared Errors," in *International Conference on Industrial Application Engineering* , Thailand, 2015.

[117] C. Legány, S. Juhász and A. Babos, "Cluster validity measurement techniques," Madrid, 2006.

[118] E. Rendón, I. M. Abundez, C. Gutierrez, S. D. Zagal, A. Arizmendi and E. M. Quiroz, "A comparison of internal and external cluster validation indexes," *Applications of Mathematics and Computer Engineering,* pp. 158-163, 2011.

[119] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 1, no. 2, p. 224–227, 1979.

[120] J. Zhong, M. Bollen and S. Rönnberg, "Towards a 100% renewable energy electricity generation system in Sweden," *Renewable Energy,* vol. 171, pp. 812-824, 2021.

[121] X. Yue, N. Patankar, J. Decarolis, A. Chiodi, F. Rogan, J. P. Deane and B. O'Gallachoir, "Least cost energy system pathways towards 100% renewable energy in Ireland by 2050," *Energy,* vol. 207, 2020.

[122] D. Icaza, D. Borge-Diez and S. P. Galindo, "Proposal of 100% renewable energy production for the City of Cuenca- Ecuador by 2050," *Renewable Energy,* vol. 170, pp. 1324-1341, 2021.

[123] T. da Luz and P. Moura, "Power generation expansion planning with complementarity between renewable sources and regions for 100% renewable energy systems," *International Transactions on Electrical Energy Systems,* pp. 1-19, 2019.

[124] A. S. Brouwer, M. Van Den Broek, A. Seebregts and A. Faaij, " Impacts of large-scale intermittent renewable energy sources on electricity systems, and how these can be modeled," *Renewable and Sustainable Energy Reviews,* vol. 33, pp. 443-466, 2014.

[125] S. Montoya-Bueno, J. I. Munoz-Hernandez and J. Contreras, "Uncertainty management of renewable distributed generation," *Journal of Cleaner Production,* vol. 138, pp. 103-118, 2016.

[126] M. R. Elkadeem, A. Younes, S. W. Sharshir, P. Campana and S. Wang, "Sustainable siting and design optimization of hybrid renewable energy system: A geospatial multi-criteria analysis," *Applied Energy,* vol. 295, 2021.

[127] M. Pluta, A. Wyrwa, W. Suwała, J. Zysk, M. Raczynski and S. Tokarski, "A Generalized Unit Commitment and Economic Dispatch Approach for Analysing the Polish Power System under High Renewable Penetration," *Energies,* vol. 13, no. 1952, pp. 1-18, 2020.

[128] B. Priyadharshini, V. Ganapathy and . P. Sudhakara, "An Optimal Model to Meet the Hourly Peak Demands of a Specific Region With Solar, Wind, and Grid Supplies," *IEEE Access,* vol. 8, pp. 13179-13193, 2020.

[129] J. H. Slusarewicz and D. S. Cohan, "Assessing solar and wind complementarity in Texas," *Renewables: Wind, Water, and Solar,* vol. 5, no. 7, pp. 1-13, 2018.

[130] C. T. M. Clack, Y. Xie and A. E. MacDonald, "Linear programming techniques for developing an optimal electrical system including high-voltage direct-current transmission and storage," *Electrical Power and Energy Systems,* vol. 68, pp. 103-114, 2015.

[131] M. R. Shaner, S. J. Davis, N. S. Lewis and K. Caldeira, "Geophysical constraints on the reliability of solar and wind power in the United States," *Energy & Environmental Science,* vol. 11, pp. 914--925, 2018.

[132] T. Nikolakakis and V. Fthenakis, "The optimummixofelectricityfromwind-andsolar-sourcesinconventional power systems:EvaluatingthecaseforNewYorkState," *Energy Policy,* vol. 39, pp. 6972-6980, 2011.

[133] J. B. Nunes, N. Mahmoudi, T. K. Saha and D. Chattopadhyay, "A stochastic integrated planning of electricity and natural gas networks for Queensland, Australia considering high renewable penetration," *Energy,* vol. 153, pp. 539-553, 2018.

[134] A. A. Prasad, R. A. Taylor and M. Kay, "Assessment of solar and wind resource synergy in Australia," *Applied Energy,* vol. 190, p. 354–367, 2017.

[135] W. Zappa and M. van den Broek, "Analysing the potential of integrating wind and solar power in Europe using spatial optimisation under various scenarios," *Renewable and Sustainable Energy Reviews,* vol. 94, pp. 1192-1216, 2018.

[136] T. Mareda, L. Gaudard and F. Romerio, "A parametric genetic algorithm approach to assess complementary options of large scale windsolar coupling," *IEEE/CAA Journal of Automatica Sinica,* vol. 4, no. 2, pp. 260-272, 2017.

[137] E. S. Takle and R. H. Shaw, "Complimentary nature of wind and solar energy at a continental mid-latitude station," *International Journal of Energy Research,* vol. 3, pp. 103-112, 1979.

[138] J. Widén, "Correlations Between Large-Scale Solar and Wind Power in a Future Scenario for Sweden," *IEEE Transactions on Sustainable Energy,* vol. 2, no. 2, pp. 177-184, 2011.

[139] F. J. Santos-Alamillos, D. Pozo-Vazquez, J. A. Ruiz-Arias, L. Von Bremen and J. Tovar-Pescador, "Combining wind farms with concentrating solar plants to providestable renewable power," *Renewable Energy,* vol. 76, pp. 539-550, 2015.

[140] S. Han, L. Zhang, Y. Liu, H. Zhang, J. Yan, L. Li, X. Lei and X. Wang, "Quantitative evaluation method for the complementarity of wind–solar–hydro power and optimization of wind–solar ratio," *Applied Energy,* vol. 236, pp. 973-984, 2019.

[141] G. Ren, J. Wan, J. Liu and D. Yu, "Spatial and temporal assessments of complementarity for renewable energy resources in China," *Energy,* vol. 177, pp. 262-275, 2019.

[142] H. Zhang, Y. Cao, Y. Zhang and V. Terzija, "Quantitative synergy assessment of regional wind-solar energy resources based on MERRA reanalysis data," *Applied Energy,* vol. 216, pp. 172-182, 2018.

[143] N. S. Thomaidis, F. J. Santos-Alamillos, D. Pozo-Vázquez and j. Usaola-García, "Optimal management of wind and solar energy resources," *Computers & Operations Research,* vol. 66, pp. 284-291, 2015.

[144] M. Koivisto, N. Cutululis and J. Ekstrom, "Minimizing Variance in Variable Renewable Energy Generation in Northern Europe," in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Boise, ID, USA, 2018.

[145] S. Jerez, R. M. Trigo, A. Sarsa, R. Lorente-Plazas, D. Pozo-Vázquez and J. P. Montávez, "Spatio-temporal complementarity between solar and wind power in the Iberian Peninsula," *Energy Procedia,* vol. 40, pp. 48-47, 2013.

[146] G. Ren, J. Lui and D. Yu, "Spatial and temporal assessments of complementarity for renewable energy resources in China," *Energy,* vol. 177, pp. 262-275, 2019.

[147] L. Stoyanov, G. Notton, V. Laza and E. M, "Wind and solar energies production complementarity for various bulgarian sites," *Revue des Energies Renouvelables SMEE'10 Bou Ismail Tipaza,* pp. 311-325, 2010.

[148] D. Heide, M. Greiner, L. von Bremen and C. Hoffmann, "Reduced storage and balancing needs in a fully renewable European power system with excess wind and solar power generation," *Renewable Energy,* vol. 36, pp. 2515-2523, 2011.

[149] J. Jurasz and J. Mikulik, "Site selection for wind and solar parks based on resources temporal and spatial complementarity – mathematical modelling approach," 2017.

[150] C. L. Archer and M. Z. Jacobson, "Supplying baseload power and reducing transmission requirements by interconnecting wind farms," *Journal of Applied Meteorology and Climatology,* vol. 46, p. 1701–1717, 2007.

[151] R. Hemmati, S. M. S. Ghiasi and A. Entezariharsini, "Power fluctuation smoothing and loss reduction in grid integrated with thermal-wind-solar-storage units," *Energy,* vol. 152, pp. 759-769, 2018.

[152] A. Coutinho, S. G. Relva, D. S. Ramos and M. E. M. Udaeta, "Hybrid Power Supply Assessment in Long Term Basis Considering Complementarity of Wind and Solar Resources," in *2019 International Conference on Smart Energy Systems and Technologies (SEST)*, Portugal, 2019.

[153] L. Ramirez Camargo, R. Zink and W. Dorner, "Spatiotemporal Modeling for Assessing Complementarity of Renewable Energy Sources in Distributed Energy Systems," *ISPRS Annals*

of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 2, no. 4, pp. 147-154, 2015.

[154] B. Jayapalan, M. Krishnan, K. Kandasamy and K. Subramanian, "Renewable Energy Penetration and Its Impact on Reliability—A Case Study of Tamil Nadu," *Journal of Computational and Theoretical Nanoscience,* vol. 14, pp. 4036-4044, 2017.

[155] R. Quijano, S. B. Botero and J. B. Domınguez, "MODERGIS application :Integrated simulation platform to promote and develop renewable sustainable energy plans, Colombian case study," *Renewable and Sustainable Energy Reviews,* vol. 16, pp. 5176-5187, 2012.

[156] J. H. Jo, M. R. Aldeman and D. G. Loomis, "Optimum penetration of regional utility-scale renewable energy systems," *Renewable Energy,* vol. 118, pp. 328-334, 2018.

[157] M. Khalid, U. Akram and S. Shafiq, "Optimal Planning of Multiple Distributed Generating Units and Storage in Active Distribution Networks," *IEEE Access,* vol. 6, pp. 55234-55244, 2018.

[158] D. Horst, M. Jentsch, M. Pfennig, I. Mitra and S. Bofinger, "Impact of renewable energies on the indian power system: Energy meteorological influences and case study of eefects on existing power fleet for rajasthan state," *Energy Policy,* vol. 122, pp. 486-498, 2018.

[159] J. Jurasz, J. Mikulik, P. Dabek, M. Guezgouz and B. Kamierczak, "Complementarity and 'resource droughts' of solar and wind energy in poland: An era5-based analysis," *Energies,* vol. 14, no. 4, 2021.

[160] C. Rosa, E. da Silva Christo, K. A. Costa and L. dos Santos, "Assessing complementarity and optimising the combination ofintermittent renewable energy sources using ground measurements," *Journal of Cleaner Production,* vol. 258, pp. 1-14, 2020.

[161] J. F. Torres, L. Ekonomou and P. Karampelas, "The Correlation Between Renewable Generation and Electricity Demand: A Case Study of Portugal," in *Energy Systems*, Berlin, Springer, 2016.

[162] Y. Li, V. G. Agelidis and Y. Shrivastava, "Wind-solar resource complementarity and its combined correlation with electricity load demand," in *2009 4th IEEE Conference on Industrial Electronics and Applications*, China, 2009.

[163] F. Canales, J. Jurasz, A. Beluco and A. Kies, "Assessing temporal complementarity between three variable energy sources through correlation and compromise programming," *Energy,* vol. 192, 2020.

[164] T. Guozden, J. P. Carbajal, E. Bianchi and A. Solarte, "Optimized Balance Between Electricity Load and Wind-Solar Energy Production," *Frontiers in Energy Research,* vol. 8, no. 16, pp. 1-10, 2020.

[165] A. A. Solomon, D. M. Kammen and D. Callaway, "The role of large-scale energy storage design and dispatch in the powergrid: A study of very high grid penetration of variable renewable resources," *Applied Energy,* vol. 134, pp. 75-89, 2014.

[166] S. Simoes, M. Zeyringer, D. Mayr, T. Huld, W. Nijs and J. Schmidt, "Impact of different levels of geographical disaggregation of wind and PV electricity generation in large energy system models: A case study for Austria," *Renewable Energy,* vol. 105, pp. 183-198, 2017.

[167] K. Iwamura and R. Kobayashi, "A Combined Geospatial Approach to Extension Planning of Wind Farms and Transmission Networks," in *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)* , Sarajevo, Bosnia and Herzegovina, 2018.

[168] Y. Schadler, V. Renken, M. Sorg , . L. Gerdes, G. Gerdes and A. Fischer, "Power transport needs for the German power grid for a major demand coverage by wind and solar power," *Energy Strategy Reviews,* vol. 34, 2021.

[169] H. M. Tróndheim, B. A. Niclasen, T. Nielsen, F. F. da Silva and C. L. Bak, "100% Sustainable Electricity in the Faroe Islands: Expansion Planning through Economic Optimisation," *IEEE Open Access Journal of Power and Energy,* vol. 8, pp. 23 - 34, 2021.

[170] G. He, A. Avrin, J. H. Nelson, J. Johnston, A. Mileva, J. Tian and D. M. Kammen, "SWITCH-China: A Systems Approach to Decarbonizing China's Power System," *Environmental Science & Technology,* vol. 50, pp. 5467-5473, 2016.

[171] M. Lenzen, B. McBain, T. Trainer, S. Jütte, O. Rey-Lescure and J. Huang, "Simulating low-carbon electricity supply for Australia," *Applied Energy,* vol. 179, pp. 553-564, 2016.

[172] L. W. Oliveira and T. C. J. Maria, "Planning of Renewable Generation in Distribution Systems Considering Daily Operating Periods," *IEEE Latin America Transactions ,* vol. 15, pp. 901 - 907, 2017.

[173] R. A. R. Candia, S. L. Balderrama, S. L. B. Subieta, J. G. P. Balderrama, V. S. Miquélez, H. J. Florero and S. Quoilin, "Techno-economic assessment of high variable renewable energy penetration in the Bolivian interconnected electric system," *International Journal of Sustainable Energy Planning and Management,* vol. 22, pp. 17-38, 2019.

[174] H. Certinay, F. A. Kuipers and A. N. Guven, "Optimal siting and sizing of wind farms," *Renewable Energy,* vol. 101, pp. 51-58, 2017.

[175] A. Vinela and E. Mortaz, "Optimal pooling of renewable energy sources with a risk-averse approach: Implications for US energy portfolio," *Energy Policy,* vol. 132, pp. 928-939, 2019.

[176] F. Monforti, T. Huld, K. Bódis, L. Vitali, M. D'Isidoro and R. Lacal-Arántegui, "Assessing complementarity of wind and solar resources for energyproduction in Italy. A Monte Carlo approach," *Renewable Energy,* vol. 63, pp. 576-586, 2014.

[177] D. Schindler, C. Jung and H. D. Behr, "On the spatiotemporal variability and potential of complementarity of wind and solar resources," *Energy Conversion and Management,* vol. 218, 2020.

[178] S. Jerez, F. Thais, I. Tobin, M. Wild, A. Colette, P. Yiou and R. Vautard, "The CLIMIX model: A tool to create and evaluate spatially-resolved scenarios of photovoltaic and wind power development," *Renewable and Sustainable Energy Reviews,* vol. 42, pp. 1-15, 2015.

[179] K. Siala and M. Y. Mahfouz, "Impact of the choice of regions on energy system models," *Energy Strategy Reviews,* vol. 25, pp. 75-85, 2019.

[180] C. Y. van Staden , H. Vermeulen and M. Groch , "Time-of-Use Feature Based Clustering of Spatiotemporal Wind Power Profiles," *Energy,* 2021.

[181] S. Satimburwa, "Transmission generation connection capacity assessment of the 2023 transmission network (GCCA – 2023)," Eskom Transmission Division, Sandton, 2021.

[182] A. Vinela and E. Mortazb, "Optimal pooling of renewable energy sources with a risk-averse approach: Implications for US energy portfolio," *Energy Policy,* vol. 132, pp. 928-939, 2019.

[183] P. Punia and M. Kaur, "Various Genetic Approaches for Solving Single and Multi-Objective Optimization Problems: A Review," *International Journal of Advanced Research in Computer Science and Software Engineering,* vol. 3, no. 7, pp. 1014-1020, 2013.

[184] M. Ming, R. Wang, Y. Zha and T. Zhang, "Multi-Objective Optimization of Hybrid Renewable Energy System Using an Enhanced Multi-Objective Evolutionary Algorithm," *Energies,* vol. 10, no. 674, pp. 1-15, 2017.

[185] Z. Liu, Z. Zhang, R. Zhuo and X. Wang, "Optimal operation of independent regional power grid with multiple wind-solar-hydro-battery power," *Applied Energy,* vol. 235, pp. 1541-1550, 2019.

[186] K. Sopian, A. Zaharim, Y. Ali, Z. M. Nopiah, J. A. Razak and N. S. Muhammad, "Optimal Operational Strategy for Hybrid Renewable Energy System Using Genetic Algorithms," *WSEAS Transactions on Mathematics,* vol. 4, no. 7, pp. 130-140, 2008.

[187] M. J. Ko, Y. S. Kim, M. H. Chung and H. C. Jeon, "Multi-Objective Optimization Design for a Hybrid Energy System Using the Genetic Algorithm," *Energies,* vol. 8, pp. 2924-2949, 2015.

[188] M. S. Ismail, M. Moghavvemi and T. M. I. Mahlia, "Genetic algorithm based optimization on modeling and design of hybrid renewable energy systems," *Energy Conversion and Management,* vol. 85, pp. 120-130, 2014.

[189] S. Katoch, S. S. Chauhan and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications,* vol. 80, p. 8091–8126, 2020.

[190] S. Thede, "An introduction to genetic algorithms," *Journal of Computing Sciences in Colleges,* vol. 20, no. 1, pp. 115-123, 2004.

[191] K. S. Tang, K. F. Man, S. Kwong and Q. He, "Genetic algorithms and their applications," *IEEE Signal Processing Magazine,* vol. 13, no. 6, pp. 22 - 37, 1996.

[192] E. Keogh and M. Pazzani, "An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback,," *4th International Conferencef Knowledge Discovery and Data Mining,* p. 239–241, 1998.

[193] N. Holjevac, T. Baškarad, J. Đakovi´, M. Krpan, M. Zidar and I. Kuzle, "Challenges of High Renewable Energy Sources Integration in Power Systems—The Case of Croatia," *Energies,* vol. 14, 2021.