



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

GENERACIÓN DE SERIES SINTÉTICAS DE RECURSOS RENOVABLES VARIABLES PARA ESTUDIOS DE OPERACIÓN Y PLANIFICACIÓN DE SISTEMAS ELÉCTRICOS

**TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN
CIENCIAS DE LA INGENIERÍA
MENCIÓN ELÉCTRICA**

JORGE LUIS VEGA HERRERA

**PROFESOR GUÍA:
RODRIGO PALMA-BEHNKE**

**PROFESOR CO-GUÍA:
FELIPE VALENCIA ARROYAVE**

**MIEMBROS DE LA COMISIÓN:
MARCELO CORTÉS CARMONA
CLAUDIA RAHMANN ZÚÑIGA**

**SANTIAGO DE CHILE
2016**

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS DE LA ING. MENCIÓN
ELÉCTRICA
POR: JORGE LUIS VEGA HERRERA
PROF. GUÍA: RODRIGO PALMA-BEHNKE
PROF. CO-GRUÍA: FELIPE VALENCIA ARROYAVE
FECHA: JUNIO 2016

GENERACIÓN DE SERIES SINTÉTICAS DE RECURSOS RENOVABLES VARIABLES PARA ESTUDIOS DE OPERACIÓN Y PLANIFICACIÓN DE SISTEMAS ELÉCTRICOS

En los últimos años a nivel mundial existe un gran interés en promover el desarrollo y penetración de energías renovables variables (ERV), tales como la energía eólica y solar fotovoltaica. El incremento de las ERV en los sistemas eléctricos de potencia (SEP) genera grandes desafíos y cambios en los paradigmas de cómo planificar la operación e inversión en los SEP debido a la variabilidad e incertidumbre de estos recursos.

Para modelar la variabilidad e incertidumbre de las ERV en los problemas de planificación de la operación y expansión de SEP se necesitan mediciones de las ERV de una longitud igual o mayor al horizonte de evaluación o una muestra, tal que permita caracterizar adecuadamente su comportamiento. En la práctica no siempre es posible tener acceso a dicha información. Debido a ello se han desarrollado diferentes metodologías y modelos para generar series sintéticas de ERV. Dichas series deben ser capaces de caracterizar la función de densidad de probabilidad, dependencia temporal y espacial de las mediciones reales. Entre los modelos más utilizados se encuentran las cadenas de Markov, procesos ARMA (del inglés AutoRegressive Moving Average) y VAR (del inglés Vector AutoRegressive).

En esta tesis se realiza una revisión bibliográfica de las metodologías y modelos desarrollados en la literatura para generar series sintéticas por tipo de recurso y tipo de dependencia que consideran en su formulación. En este contexto, el objetivo general de este trabajo es conocer y sistematizar las diferentes propuestas en la literatura para la generación de series sintéticas de velocidad de viento, potencia eólica y radiación. Además, definir y proponer criterios para la adecuada selección de modelos a ser utilizados en operación y planificación de sistemas eléctricos de potencia. Para ello se realiza una propuesta metodológica que consta de seis etapas, en donde se describe detalladamente los pasos a seguir así como los análisis estadísticos necesarios para elegir adecuadamente un modelo para generar series sintéticas.

La metodología es aplicada a problemas de planificación de operación de mediano y muy corto plazo. Se observa que los modelos VAR son capaces de caracterizar los estadísticos totales, mensuales y estructuras de dependencia de las ERV cuando se utilizan procesamientos y transformaciones adecuadas. Además, se compara con otros modelos tales como CARMA (del inglés Contemporaneous AutoRegressive Moving Average), PCA (del inglés Principal Component Analysis) + ARMA, VAR con distintos procesamientos y teoría de Cópula, usando la familia de cópulas gaussianas para esta última.

Se concluye que la metodología propuesta es capaz de sistematizar el proceso de selección de modelos temporales y espacio-temporales en función del problema de SEP y las características estadísticas de las series de tiempo asociadas a las ERV. Asimismo, en función del estado del arte se sistematiza el proceso para definir transformaciones, estimar el orden y ajustar adecuadamente los modelos VAR. Finalmente, se observa la relevancia de verificar el comportamiento estadístico de las series de tiempo. Si no se realiza un análisis adecuado y se aplican los modelos desarrollados en la literatura sin verificar sus supuestos y condiciones de aplicación, se podrían utilizar modelos que no caractericen adecuadamente ni las contribuciones energéticas estacionales ni las estructuras de dependencias.

Como principal trabajo futuro se plantea desarrollar una herramienta computacional basada en un sistema experto que permite automatizar la toma de decisión para personas que son expertas en SEP, pero no en modelos para generar series sintéticas, ayudando a mejorar las prácticas actuales referentes a la caracterización de la variabilidad e incertidumbre de las ERV a través de las series sintéticas.

Agradecimientos

Me gustaría agradecer a Dios, en primer lugar por haberme dado la oportunidad de entrar a la Universidad de Chile y realizar mis estudios de Magíster. En segundo lugar por darme ánimo cada año para estudiar y tener el deseo de ser una mejor persona y profesional. Sin su ayuda no hubiera podido lograrlo.

A mi esposa, por su paciencia y ayuda en esta etapa, sin sus revisiones de la tesis no hubiera podido terminar. Gracias por ayudarme día a día a enfocarme y cumplir con nuestras metas.

A mis viejitos, Jorge Vega Palma y Renata Herrera, que siempre han estado ahí cuando lo he necesitado, su apoyo incondicional ha bendecido mi vida.

A mi profesor guía, Rodrigo Palma, por desafiarme constantemente a buscar la excelencia en las tareas realizadas, por corregirme y compartir de su experiencia y conocimientos, han sido una gran ayuda en mi proceso de formación como investigador.

A mi profesor co-guía, amigo y jefe jefe..., Felipe Valencia, por escucharme con paciencia mis inquietudes y ayudarme a cumplir con los objetivos de este trabajo de tesis.

A mis amigos de postgrado, Diego Ortiz, Jacqueline Llanos y Oscar Núñez, por ayuda y apoyo incondicional. Gracias por esta allí cuando lo necesité y por brindarme de su amistad.

Y por último y no menos importante a la Universidad de Antofagasta y al programa de becas CONICYT-PCHA/Magíster Nacional/2013-22130405 los cuales me han brindado su ayuda y confianza para perfeccionarme y ser un aporte a la región de Antofagasta

Tabla de contenido

Índice de figuras.....	v
Índice de tablas.....	ix
1 Introducción.....	1
1.1 Motivación	1
1.2 Hipótesis de la tesis	2
1.3 Objetivos de la tesis.....	2
1.3.1 Objetivo general	2
1.3.2 Objetivos específicos.....	2
1.4 Alcances	2
1.5 Estructura de la tesis.....	2
2 Estado del arte	4
2.1 Introducción	4
2.2 Evolución de energía solar fotovoltaica y eólica a nivel mundial.....	4
2.3 Problemas de operación y planificación en sistemas eléctricos	6
2.4 Impacto en la energía solar y eólica	9
2.4.1 Características	9
2.4.2 Impacto en la operación de sistemas eléctricos.....	11
2.4.3 Impactos en la planificación de sistemas eléctricos	13
2.5 Generación de series sintéticas.....	14
2.5.1 Modelos para generar series sintéticas que consideran dependencia temporal	14
2.5.2 Modelos para generar series sintéticas que consideran dependencia temporal y espacial ..	21
2.6 Análisis de modelos de generación de series sintéticas y problemas detectados	24
2.7 Resumen y conclusiones del estado del arte	25
3 Propuesta metodológica para selección de modelos de generación de series sintéticas.....	31
3.1 Estructura general de la metodología propuesta.....	31
3.2 Descripción de las componentes de la metodología propuesta	32
3.2.1 Definir problema de SEP.....	32
3.2.2 Definir tipo de serie del recurso solar y/o eólico disponible	36
3.2.3 Etapa de decisión: ¿Existe información necesaria para desarrollar el estudio?.....	38
3.2.4 Análisis estadístico de las series del recurso solar y/o eólico.....	38
3.2.5 Definir conjunto de modelos candidatos	45
3.2.6 Selección del modelo	47
3.3 Implementación.....	57
3.4 Resumen.....	59
4 Casos de estudio y análisis de resultados	60
4.1 Introducción	60

4.2	Primer caso de estudio: Modelo de planificación de la operación de mediano plazo	60
4.2.1	Resultados gráficos del modelo seleccionado	72
4.3	Segundo caso de estudio: Planificación de la operación de muy corto plazo	78
4.3.1	Resultados gráficos del modelo seleccionado	88
5	Conclusiones	93
5.1	Conclusiones respecto al trabajo de tesis	93
5.2	Trabajos futuros.....	94
6	Bibliografía.....	95
7	Anexo 1: Fundamentos matemáticos.....	104
7.1	Procesos estocásticos estacionarios[123].....	104
7.1	Transformada integral [78].....	105
7.2	Dependencia entre variables aleatorias	106
7.2.1	Independencia [136].....	106
7.2.2	Medidas de dependencia [136][3].....	106
7.3	Muestreo multivariable [137].....	110
7.4	Criterios AIC.....	110
8	Anexo 2: Modelos que consideran dependencia temporal	112
8.1	Antecedentes generales del anexo.....	112
8.2	Cadenas de Markov [50], [51],[53].....	112
8.3	ARMA.....	115
9	Anexo 3: Modelos espaciotemporales.....	119
9.1	Antecedentes generales del anexo.....	119
9.2	VAR [126], [131][138]	119
9.3	CARMA [7], [100], [107], [108].....	125
9.4	PCA+ARMA [139].....	128
9.5	Modelos teoría de Cópula Gaussiana [128]	132
10	Anexo 4: Listado de modelos para generar series sintéticas	138
11	Anexo 5: Pruebas de hipótesis para autocorrelaciones.....	142
11.1	Modelo univariable [126].....	142
11.2	Modelo multivariable [126]	142

Índice de figuras

Figura 2.1. Evolución en capacidad instalada de ERNC [15]	5
Figura 2.2. Evolución capacidad instalada de ERV en Chile.....	6
Figura 2.3. Modelos de planificación y operación en función del detalle y escala de tiempo considerados [28].....	8
Figura 2.4. Función de distribución acumulada (entre los percentiles 95% y 100%) de cambios en radiación y potencia fotovoltaica para varios periodos de tiempo durante 1 día para un sistema de 30 kW fotovoltaicos (izquierda) y un sistema múltiple del orden de los MW fotovoltaicos (derecha) [32].....	10
Figura 2.5. Coeficientes de correlación para cambios en el índice de claridad para diferentes distancias entre sitios y diferentes intervalos de tiempo [32].	11
Figura 2.6. Ciclo diario de velocidad de viento a 90 metros. Los puntos negros indican el promedio de viento en cada hora del día. Las líneas rojas indican el valor mediano de las distribuciones horarias. Las barras azules representan el rango intercuartil. Las barras negras indican el rango de los valores horarios excluyendo valores extremos.	11
Figura 2.7. Gráfica de la demanda del sistema (azul) y la demanda neta (verde) en donde se ejemplifica distintos impactos en los SE debido a la integración de ERNC[28].	12
Figura 2.8 Número de horas donde las unidades de carbón y de ciclo combinado aumentaron o disminuyeron severamente su generación [42]	13
Figura 2.9. Distribución de frecuencia de la potencia del aerogenerador Vestas V126-3.0 MW a 90 metros. Las barras azules representan el porcentaje de valores dentro de cada intervalo. La línea roja es la distribución de frecuencia acumulada	16
Figura 2.10. Comparación de las funciones de autocorrelación simple entre la serie real y la serie sintética generada utilizando cadenas de Markov de 1° y 2° orden reportado en [50].....	17
Figura 2.11. Comparación de las funciones de autocorrelación simple entre la serie real y la serie sintética generada utilizando cadenas de Markov de 1° reportado en [61]	17
Figura 2.12. Esquema de modelos para generar series sintéticas de velocidad de viento y potencia eólica.....	26
Figura 2.13. Esquema de modelos para generar series sintéticas de radiación y potencia fotovoltaica.....	27
Figura 2.14. Esquema de modelos para generar series sintéticas que consideran recurso solar y eólico.....	27
Figura 2.15. Esquema de modelos para generar series sintéticas que consideran como variable radiación, velocidad de viento y demanda	27
Figura 2.16. Esquema de problemas de SEP y sus respectivos horizontes de tiempo	27
Figura 2.17. Esquema con distintas variantes de modelos VAR reportados en la literatura.....	28
Figura 2.18. Procesamientos previos realizados para remover tendencia, estacionalidad y transformar las series de ERV a normales.....	28
Figura 3.1. Diagrama de flujo general de metodología propuesta	32
Figura 3.2. Mapa conceptual etapa 1: Definir problema de SEP	35
Figura 3.3. Relación entre preguntas de la etapa 1 con el resto de las etapas.	36
Figura 3.4. Esquema general etapa 2 de la metodología propuesta.....	36
Figura 3.5. Detalle de subsección Mediciones disponibles en lugares de interés	37
Figura 3.6. Detalle de subsección modelos fenomenológicos.....	37
Figura 3.7. Relación entre preguntas de la etapa 2 con el resto de las etapas de la metodología propuesta.	38
Figura 3.8. Mapa conceptual de la etapa 4: Análisis estadístico de las series del recurso solar y/o eólico.....	39
Figura 3.9. Diagrama de cajas de velocidad de viento en zona de Antofagasta (Datos obtenidos en [17]).....	41
Figura 3.10. Diagrama de cajas de velocidad de viento mensual en Nueva Zelanda (Datos obtenidos en [125]).....	41
Figura 3.11. Ejemplo de función de autocorrelación de serie de potencia eólica agregada en las zonas de Antofagasta a Coquimbo.....	42
Figura 3.12. Función de autocorrelación cruzada de dos series de Nueva Zelanda (datos obtenidos de [125])	43

Figura 3.13. Resultado de prueba de hipótesis Ljung-Box para dos series correlacionadas espacialmente	43
Figura 3.14. Diagrama de flujo de análisis estadístico	44
Figura 3.15. Relación entre preguntas de la etapa 4 con el resto de las etapas de la metodología propuesta	45
Figura 3.16. Mapa conceptual etapa 4: Definir conjunto de modelos candidatos	46
Figura 3.17. Diagrama de proceso para seleccionar modelos candidatos	48
Figura 3.18. Mapa conceptual etapa 6: Selección de modelo	49
Figura 3.19. Subetapa 1: determinación de procesamiento previo	52
Figura 3.20. Subetapa 2: determinación de parámetros de los modelos	53
Figura 3.21. Diagrama de flujo para selección y ajuste modelos VAR(p)	55
Figura 3.22. Subetapa 3: generación de series sintéticas	55
Figura 3.23. Subetapa 4: cálculo de indicadores para evaluar bondad de los modelos-esquema general	56
Figura 3.24. Subetapa 5: Seleccionar modelo corregir modelo	57
Figura 3.25. Diagrama de flujo de implementación de la metodología	58
Figura 4.1. Ubicación de mediciones en Nueva Zelanda [125]	61
Figura 4.2. Histogramas de las series de velocidad de viento en los sitios STH1 y CKS1	62
Figura 4.3. Series de tiempo de tiempo de velocidades de viento para los sitios STH1 y CKS1	63
Figura 4.4. Diagrama de cajas para variaciones horarias y mensuales para los sitios STH1 y CKS1	64
Figura 4.5. Funciones de autocorrelación simple y parcial correspondientes a los sitios STH1 y CKS1	65
Figura 4.6. Gráfico de p-valores de la prueba de hipótesis Ljung box para modelos multivariantes	65
Figura 4.7. Diagrama de dispersión entre los sitios STH1 y CKS1	66
Figura 4.8. Función de autocorrelación cruzada	66
Figura 4.9. RMSRE para CCF y estadísticos mensuales	71
Figura 4.10. RMSRE promedio ponderado	72
Figura 4.11. RMSRE aplicado sobre estadísticos total al sitio STH1	72
Figura 4.12. Comparación de histograma normalizado para sitio STH1	73
Figura 4.13. Comparación de histograma normalizado para sitio CKS1	74
Figura 4.14. Comparación de ACF y PACF para el sitio STH1	74
Figura 4.15. Comparación de ACF y PACF para el sitio STH1	74
Figura 4.16. Comparación de diagramas de dispersión	75
Figura 4.17. Comparación de función de autocorrelación cruzada	75
Figura 4.18. Comparación de diagramas de dispersión para el sitio STH1	76
Figura 4.19. Comparación de diagramas de dispersión para el sitio CKS1	76
Figura 4.20. Comparación de diagrama de cajas horarios asociados al sitio STH1	77
Figura 4.21. Comparación de diagrama de cajas horarios asociados al sitio CKS1	77
Figura 4.22. Comparación de diagrama de cajas mensuales asociados al sitio STH1	78
Figura 4.23. Comparación de diagrama de cajas mensuales asociados al sitio CKS1	78
Figura 4.24. Histograma de series de potencia fotovoltaica correspondiente a la región de Atacama (Datos normalizados)	81
Figura 4.25. Series de tiempo de tiempo de potencia fotovoltaica y eólica en las regiones de Atacama y Coquimbo respectivamente	82
Figura 4.26. Diagrama de cajas para variaciones horarias de series de potencia fotovoltaica y eólica en las regiones de Atacama y Coquimbo	83
Figura 4.27. Funciones de autocorrelación simple y parcial correspondientes a los sitios PV-Atacama y WT-Coquimbo	84
Figura 4.28. Gráfico de p-valores de la prueba de hipótesis Ljung box para modelos multivariantes aplicados a las series de potencia fotovoltaica y eólica	85
Figura 4.29. Diagrama de dispersión entre sitios PV-Atacama y WT-Coquimbo	85
Figura 4.30. Función de autocorrelación cruzada	85
Figura 4.31. Gráfica con aplicación de prueba de hipótesis Ljung-Box para verificar que los residuos no se encuentren correlacionados	87

Figura 4.32. Comparación de histogramas en sitio WT-Coquimbo.....	88
Figura 4.33. Comparación de histogramas en sitio PV-Atacama.....	88
Figura 4.34. Diagrama de cajas en sitio PV-Atacama.....	89
Figura 4.35. Diagrama de cajas para sitio WT-Coquimbo.....	89
Figura 4.36. Comparación de ACF y PACF en sitio PV-Atacama.....	90
Figura 4.37. Comparación de ACF y PACF en sitio WT-Coquimbo.....	90
Figura 4.38. Comparación de CCF.....	91
Figura 4.39. Comparación de diagrama de dispersión entre sitio PV-Atacama y WT-Coquimbo.....	91
Figura 4.40. RMSRE para distintos estadísticos aplicados al modelo VAR(72).....	92
Figura 7.1. Aplicación de la transformada integral.....	106
Figura 8.1. Comparación de histogramas en sitio CKS1 utilizando cadena de Markov de primer orden..	114
Figura 8.2. Comparación de ACF y PACF en sitio CKS1 utilizando cadena de Markov de primer orden.....	115
Figura 8.3. Gráfica de indicador pare diferentes estadísticos del sitio CKS1 utilizando cadena de Markov de primer orden.....	115
Figura 8.4. ACF y PACF de los residuos del modelo ARMA(3,3) asociado al sitio CKS1.....	117
Figura 8.5. Comparación de ACF y PACF en sitio CKS1 utilizando modelo ARMA(3,3).....	117
Figura 8.6. Comparación de histogramas en sitio CKS1 utilizando modelo ARMA(3,3).....	118
Figura 8.7. Comparación de diagrama de caja para variaciones horarias en sitio CKS1 utilizando modelo ARMA(3,3).....	118
Figura 8.8. Gráfica de indicador para distintos estadísticos en sitio CKS1 utilizando modelo ARMA(3,3).....	118
Figura 9.1. Metodologías propuestas para transformar series de tiempo a su forma estandarizada.....	120
Figura 9.2. Histograma normalizado y qqplot para serie transformada asociada al sitio STH1.....	122
Figura 9.3. Resultado de aplicación prueba de hipótesis Kruskas Wallis a serie transformada asociada a sitio STH1.....	122
Figura 9.4. Diagrama de cajas sitio STH1 transformado.....	123
Figura 9.5. ACF y PACF de los residuos asociados al sitio STH1.....	124
Figura 9.6. CCF de los residuos.....	124
Figura 9.7. Gráfica de los p-valores tras aplicar la prueba de hipótesis Ljung-box.....	125
Figura 9.8. Histograma y qqplot para serie transformada del sitio STH1.....	126
Figura 9.9. Comparación de ACF y PACF entre serie real y sintética para el sitio STH1 utilizando modelo CARMA.....	127
Figura 9.10. Comparación de histogramas entre serie real y sintética para los sitios STH1 y CKS1 utilizando modelo CARMA.....	127
Figura 9.11. Gráfica con indicadores RMSRE utilizando modelo CARMA.....	128
Figura 9.12. Histograma y diagrama de dispersión para series transformadas (eje x sitio CKS1, eje y sitio STH1).....	129
Figura 9.13. Comparación de histogramas entre serie real y sintética para los sitios STH1 y CKS1 utilizando PCA+ARMA.....	130
Figura 9.14. Comparación de ACF y PACF para el sitio CKS1 utilizando PCA+ARMA.....	131
Figura 9.15. Comparación de función de autocorrelación cruzada utilizando PCA+ARMA.....	131
Figura 9.16. Función de autocorrelación cruzada de serie transformada utilizando PCA.....	132
Figura 9.17. Gráfica de RMSRE para distintos estadísticos utilizando PCA+ARMA.....	132
Figura 9.18. Matriz de correlación para 48 variables aleatorias.....	133
Figura 9.19. Funciones de distribución acumulada para cada sitio y hora.....	134
Figura 9.20. Comparación entre serie real y sintética para los sitios STH1 y CKS1 usando una cópula gaussiana.....	135
Figura 9.21. Comparación de Boxplot horaria para sitio CKS1 utilizando cópula gaussiana.....	135
Figura 9.22. Comparación de Boxplot horaria para sitio STH1 utilizando cópula gaussiana.....	136
Figura 9.23. Comparación de ACF y PACF para sitio STH1 utilizando cópula gaussiana.....	136

Figura 9.24. Gráfica de matriz de correlación utilizando cópula gaussiana.....	137
Figura 9.25. RMSRE para sitio STH1 usando cópula gaussiana	137

Índice de tablas

Tabla 2.1. Crecimiento de capacidad instalada de ERNC periodo 2012-2013[16].....	5
Tabla 2.2. Cuadro resumen del estado de proyectos ER a diciembre 2015 [18].....	6
Tabla 2.3. Descripción de problemas de operación y planificación según distintos horizontes de evaluación [28].....	9
Tabla 2.4. Resumen de modelos para generación de series sintéticas con dependencia temporal de acuerdo a la literatura (Fuente: elaboración propia).....	29
Tabla 2.5. Resumen de modelos para generación de series sintéticas con dependencia espacio-temporal de acuerdo a la literatura (Fuente: elaboración propia).....	30
Tabla 4.1. Resumen de análisis a series transformadas y residuos de los modelos propuestos	69
Tabla 4.2. Metodologías implementadas (Parte 1).....	69
Tabla 4.3. Metodologías implementadas (Parte 2).....	70
Tabla 4.4. Capacidad instalada de centrales solares fotovoltaicas y eólicas (fuente: CDEC-SIC)	80
Tabla 4.5. AIC para estimación del orden del modelo VAR(p).....	87
Tabla 8.1. Comparación de estadísticos entre serie real y sintética en sitio CKS1 utilizando un modelo de cadena de Markov de primer orden.....	114
Tabla 9.1. Criterio AIC para distintos valores de p.....	123
Tabla 10.1. Listado de modelos para generar series sintéticas reportados en la literatura parte n° 1	138
Tabla 10.2. Listado de modelos para generar series sintéticas reportados en la literatura parte n° 2	139
Tabla 10.3. Listado de modelos para generar series sintéticas reportados en la literatura parte n° 3	140
Tabla 10.4. Listado de modelos para generar series sintéticas reportados en la literatura parte n° 4	141

1 Introducción

1.1 Motivación

A nivel mundial en las últimas décadas se ha experimentado un gran incremento en la generación de electricidad en base a fuentes de energías renovables (ER). Diversos países tienen metas de participación de generación o han generado distintos esquemas de incentivos para impulsar el desarrollo de este tipo de energías. Dentro de las ER que han experimentado mayor crecimiento se encuentra la energía solar fotovoltaica y la energía eólica. Este hecho ha ocurrido debido a la madurez tecnológica y disminución sostenida en los costos de inversión, haciéndolas más competitivas [1]. La capacidad instalada de energía eólica durante el año 2010 llegó a los 200 GW con 83 proyectos en operación en 83 países y se proyecta que 1000 GW estarán en operación para el 2020. La energía solar fotovoltaica también ha experimentado un crecimiento exponencial: para el 2011 se tenía cerca de 40 GW en todo el mundo, esto equivale a más de siete veces la capacidad instalada el 2005 [2].

El incremento experimentado y la naturaleza de estas fuentes han generado grandes desafíos en la operación y planificación de sistemas eléctricos. Estas fuentes de generación producen cambios en la operación, en general la demanda neta manifiesta mayor variabilidad, lo que implica que el sistema eléctrico requiera que exista mayor variación en el encendido y apagado de las unidades generadoras convencionales y se hace necesario cambiar los requerimientos de reserva para operar en forma segura y confiable, en otras palabras se necesita mayor flexibilidad en el sistema eléctrico. Las centrales construidas en base a estas fuentes son variables e inciertas, por ello han sido denominadas en la literatura como centrales de generación no despachables, debido que el suministro de potencia está directamente relacionado con la disponibilidad del recurso primario tal como radiación y velocidad del viento [3]. Por otro lado, en la planificación de la inversión de generación y/o transmisión a medida que la capacidad instalada de energías renovables variables (ERV) empiezan a aumentar surge la necesidad de considerar mayores detalles de la operación, cambiando un modelo de curvas de duración a uno que considere la evolución cronológica de las ERV [1] [4].

Numerosos son los trabajos que han buscado resolver los problemas de operación y planificación de los sistemas eléctricos (SE) en presencia de incertidumbre y variabilidad ocasionadas por las ERV. Una de las formas en que se ha modelado la incertidumbre en los sistemas eléctricos consiste en la utilización de optimización estocástica y robusta. Estas técnicas han sido utilizadas para resolver problemas de predespacho, planificación de la generación y transmisión [5][6]. Todos los modelos ya mencionados necesitan tener series de tiempo del recurso que se está considerando para modelar adecuadamente la variabilidad e incertidumbre. Esta es una tarea compleja, ya que en la práctica frecuentemente no es posible disponer de series de tiempo de ERV cuya longitud sea igual al problema de SEP que se desea resolver o la muestra no contiene la información necesaria para analizar adecuadamente el impacto de las ERV en los SEP. Como respuesta a esta dificultad, se ha estudiado y desarrollado distintos modelos para generar series sintéticas, las cuales son utilizadas como datos de entrada para las herramientas de solución de los problemas de SEP [7].

Las series sintéticas son series que se generan a través de ecuaciones matemáticas y tienen como objetivo replicar las propiedades estadísticas de la serie original. Las variables con incertidumbre son tratadas como variables aleatorias, por lo tanto una serie sintética representa una realización de un proceso estocástico [8]. Las series sintéticas han sido utilizadas para determinar niveles de reserva, estudiar confiabilidad en los SE, generar escenarios para predespachos estocásticos, determinar niveles de almacenamiento, entre muchos otros problemas [9]-[10]. Además, en el estudio y análisis del impacto de las ERV son de gran ayuda, porque es posible generar series de tiempo, las cuales representan condiciones de operacional que podrían ocasionar consecuencias más desfavorables en la operación y planificación SE o como una herramienta para utilizar técnicas probabilísticas en la toma de decisiones [8].

Los procesos estocásticos que se desean simular en este trabajo corresponden al recurso solar y eólico dada su relevancia actual y proyección. Estos son procesos estocásticos complejos y por ello es que se han

desarrollado diversos modelos en la literatura especializada tales como: ARMA, Cadenas de Markov, teoría de Cópula, entre otros [11]. Se observa que por un lado existe un campo fructífero de desarrollo de modelos para generar series sintéticas y otro que desarrolla modelos de operación y planificación para estudiar la integración de ERV pero no existe mucha comunicación entre ambos mundos, no quedando claro en la literatura qué modelo utilizar o cuáles son las características en las que es posible utilizar un determinado modelo [12]. Debido a lo expuesto anteriormente, este trabajo busca generar un puente entre ambos campos de investigación, aportando en el análisis de los distintos modelos existentes en la literatura y en base a ello proponer una metodología que permite sistematizar la selección de modelos en función de los objetivos de los problemas de SE que se desea resolver.

1.2 Hipótesis de la tesis

La generación adecuada de series sintéticas de velocidad de viento, potencia eólica y radiación es un factor clave para un desempeño de los modelos de sistemas eléctricos de potencia.

1.3 Objetivos de la tesis

1.3.1 Objetivo general

El objetivo general de este trabajo es conocer y sistematizar las diferentes propuestas en la literatura para la generación de series sintéticas de velocidad de viento, potencia eólica, radiación solar y potencia fotovoltaica. Además, definir y proponer criterios para la adecuada selección de modelos a ser utilizados en operación y planificación de sistemas eléctricos de potencia

1.3.2 Objetivos específicos

Los objetivos específicos de este trabajo son:

- Conocer los modelos existentes para generar series sintéticas de recursos renovables propuestos en la literatura especializada
- Implementar modelos que consideren dependencia temporal y espacio-temporal de recursos renovables y validar con series de tiempo en distintos lugares geográficos
- Proponer una metodología que permita elegir modelos adecuados para estudios de sistemas eléctricos
- Validar la metodología propuesta en un caso de estudio particular

1.4 Alcances

En este trabajo se desarrollan y aplican modelos para series de velocidad de viento, potencia eólica y radiación solar. Los aspectos específicos de conversión entre el recurso primario a potencia así como aspectos fenomenológicos de la velocidad y radiación no son abordados. La metodología busca sistematizar, organizar y jerarquizar la información obtenida en el estado del arte. Sin embargo, no se considera el desarrollo de un algoritmo que desarrolle automáticamente el proceso propuesto en la metodología para la selección de los modelos.

1.5 Estructura de la tesis

En el capítulo 2 se describe el estado del arte, entregando una visión general de la evolución de las ER en el mundo y en el país así como los impactos en los SE debido a su integración. Como componente principal de este capítulo se presenta una revisión de los distintos modelos para generar series sintéticas descomponiendo en aquellos modelos que buscan caracterizar la dependencia temporal y aquellos que buscan caracterizar la dependencia espacio temporal de los recursos. Finaliza el capítulo con un resumen y principales conclusiones de la revisión.

En el capítulo 3 se detalla la propuesta metodológica para sistematizar la selección de modelos. Se analiza detalladamente cada etapa del proceso y se termina con la implementación. En la sección de validación se muestran la aplicación de algunos modelos para generar series sintéticas, verificando los indicadores y estadísticos que se consideran en la metodología.

En el capítulo 4 se presentan los casos de estudio y se discuten los resultados. Se presenta un caso ilustrativo y se aplica la metodología paso a paso sin considerar ajuste de modelos, haciendo énfasis en los criterios de selección. Un segundo caso es implementado en donde se aplica la metodología pero se muestra paso a paso los resultados estadísticos y de ajuste de los modelos.

Finalmente, en el capítulo 5 se presentan las conclusiones referentes al desarrollo de los modelos y la aplicación de la metodología propuesta. Los fundamentos matemáticos y los algoritmos para la generación de series sintéticas son presentados en los Anexos.

2 Estado del arte

2.1 Introducción

En este capítulo se presenta la evolución de las ER dando un mayor énfasis en la energía solar y eólica. Luego se discuten sus principales características desde un punto de vista general, sin considerar los fenómenos físicos asociados a la radiación solar y velocidad de viento. Al terminar de entender las características se analizan y discuten los impactos en los SE debido a la integración de grandes volúmenes de ERV. Los impactos se dividen en aquellos relacionados con la operación, planificación de la operación y planificación de la inversión.

Se describen brevemente los problemas que existen en los SE con sus respectivos horizontes de evaluación. El objetivo de esta descripción es definir los tipos de modelos básicos utilizados y ver en qué problemas se han utilizado o es posible utilizar series sintéticas en los SE, por ende no se detalla su formulación matemática. Se analizan los distintos modelos desarrollados en la literatura especializada y se clasifican según tipo de dependencia que consideran en su formulación, se discuten los problemas encontrados durante la revisión así como algunos modelos de series temporales que aún no han sido aplicados en el contexto de los SE. Para finalizar se realiza un resumen con los principales hallazgos de la revisión y se concluye en base a ellos. Esta sección es importante porque se genera una base de datos que es utilizada en la propuesta metodológica para la selección de modelos que será descrita en el siguiente capítulo.

2.2 Evolución de energía solar fotovoltaica y eólica a nivel mundial

Las energías renovables (ER) pueden ser agrupadas en fuentes despachables, tales como las centrales hidroeléctricas, geotérmicas y biomasa, y no despachables, variables o intermitentes, tales como las centrales eólicas, fotovoltaicas, concentración solar y mareomotriz. Cuando se dice que una fuente es despachable, se refiere a que es posible controlar su respuesta en base a los distintos requerimientos que posee el SE, tal como la variación de la demanda. En general una central despachable se encuentra siempre disponible (con excepción de las mantenciones) y poseen factores de planta elevados. En contraste, la generación de electricidad en base a fuentes de energía no despachables depende de las fuentes primarias intermitentes y variables [3].

El incremento e integración de ER es una componente principal en las políticas energéticas de muchos países. En la actualidad más de 140 países poseen metas de integración de ER. Por ejemplo, la Unión Europea (UE) tiene la meta de lograr que un 37% del consumo energético sea en base a renovables, lo que implica que se podría llegar a cerca del 51% -68% en el sector eléctrico [13]. Existen otros países pequeños que poseen metas del orden del 50% o aún más.

De acuerdo a la IEA [14] (del inglés “*International Energy Agency*”) para el año 2012 la generación en base a energías renovables a nivel mundial fue aproximadamente 21% (22721 TWh), con 16% de centrales hidroeléctricas, 2,3% energía eólica, 1,9% biomasa, 0,3% geotérmica y 0,4% fotovoltaica. En términos de capacidad instalada 27,8%(19% hidroeléctrica, 4,9% eólica, 1,7% biomasa, 0,2% geotérmica, 1,7% fotovoltaica) de un total de 5683 GW. Sin embargo, en los últimos años se ha experimentado un crecimiento rápido en el mundo, principalmente de energía eólica y fotovoltaica (ver Figura 2.1). Esto ha ocurrido por las políticas de incentivo y una disminución en los costos de estas tecnologías. Por ejemplo, durante los años 2009-2013 la energía solar fotovoltaica, energía solar de concentración y energía eólica crecieron a tasas anuales del 55%, 48% y 21% respectivamente. La capacidad acumulada de centrales eólicas y fotovoltaicas en el año 2011 fue de 236 GW y 70 GW respectivamente, mientras que sus valores para finales del 2013 fueron 318 GW y 138 GW. Distintas proyecciones indican que en los próximos años seguirán incrementando las ER, especialmente la energía eólica y fotovoltaica [15]. Los países que lideran la capacidad instalada de energías renovables fueron: China, Japón, Alemania, Estados Unidos e Inglaterra. Japón y Estados Unidos muestran un aumento considerable en centrales fotovoltaicas (ver Tabla 2.1).

Tabla 2.1. Crecimiento de capacidad instalada de ERNC periodo 2012-2013[16]

País	Energía eólica [%]	Energía fotovoltaica [%]	ER [%]
China	21	168	33
Japón	2	106	76
Alemania	11	11	11
Estados Unidos	2	65	9
Inglaterra	26	59	31
India	9	78	14

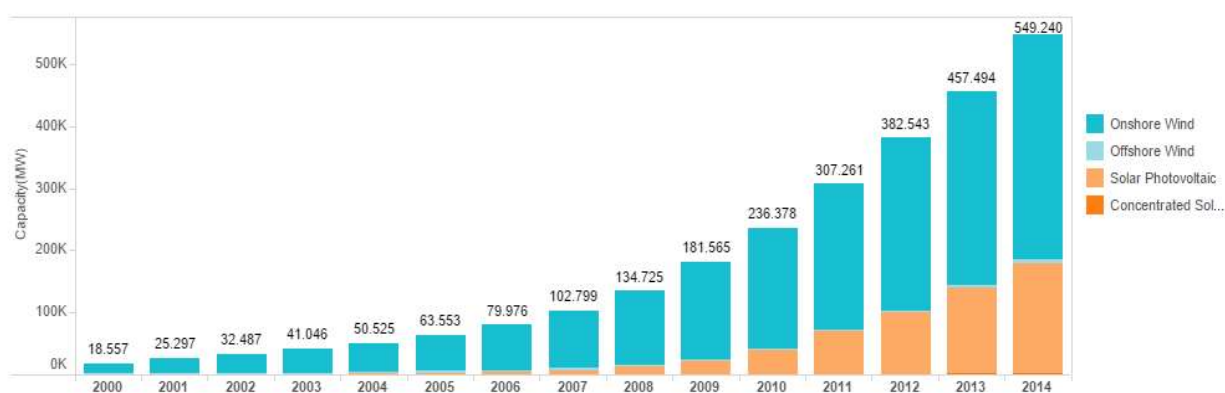


Figura 2.1. Evolución en capacidad instalada de ERNC [15]

En Chile también se ha experimentado un importante aumento de las tecnologías ER, esto debido a distintas políticas implementadas en el sector y a la disponibilidad de los recursos primarios en especial el recurso solar: Chile posee el desierto más soleado de la tierra, convirtiéndose en uno de los mejores lugares para el desarrollo de la energía solar [17]. En la Figura 2.2 se observa la evolución de la capacidad instalada de centrales eólicas y fotovoltaicas, llegando a fines del año 2015 a 901 [MW] y 750 [MW] respectivamente. El incremento entre los años 2014-2015 corresponde al 8% y 105%. Por otro lado, según informa CIFES¹ existen 10350 [MW] (ver Tabla 2.2) con resolución ambiental aprobada, esto corresponde a un 53% de la capacidad total de los sistemas SING-SIC. Este escenario de crecimiento que se está experimentando en forma internacional motiva a la búsqueda de mejores modelos para caracterizar adecuadamente las ER que son variables y no despachables.

¹ Centro para la Innovación y Fomento de las Energías Sustentables (CIFES)

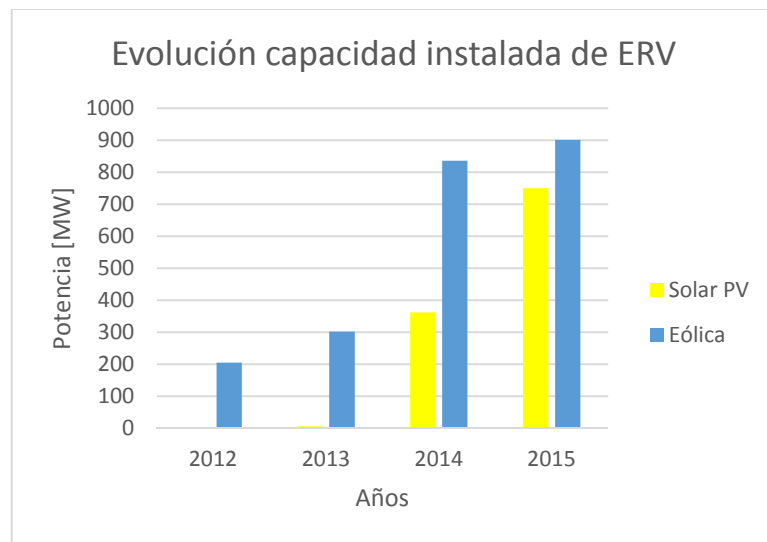


Figura 2.2. Evolución capacidad instalada de ERV en Chile

Tabla 2.2. Cuadro resumen del estado de proyectos ER a diciembre 2015 [18]

Tecnología	Operación [MW]	Construcción [MW]	RCA Aprobada [MW]	En Calificación [MW]
Biomasa	417	0	82	77
Biogás	44	0	8	0
Eólica	901	224	5820	1578
Geotermia	0	48	120	110
Mini Hidro	397	54	429	4134
Solar -PV	750	2270	10350	4134
Solar -CSP	0	110	980	
Total	2509	2706	17789	10033

2.3 Problemas de operación y planificación en sistemas eléctricos

Los sistemas eléctricos de potencia (SEP) son sistemas grandes y complejos, modelarlos adecuadamente es una tarea desafiante. Los SEP se encuentran sujetos a diferentes condiciones de operación, las cuales se encuentran constantemente cambiando. Algunas de estas condiciones pueden ser anticipadas por los encargados de ejecutar las acciones de control, permitiendo tomar buenas decisiones. Sin embargo, otras cambian sin previo aviso, requiriendo contar con estrategias de operación y control que permitan mantener una operación estable y segura [19].

Debido al tamaño y complejidad de los SEP su operación y planificación es dividida en base a diferentes ventanas de tiempo, desde los milisegundos hasta los años. Cada ventana de tiempo tiene sus propias restricciones, mecanismos de operación, requerimientos de información, implicaciones económicas, desafíos técnicos y simplificaciones. Dadas las capacidades de cómputo actuales, existe un trade-off entre el nivel de detalle capturado y el horizonte de tiempo considerado [20]. En la Figura 2.3 es posible observar cómo, a medida que la ventana de tiempo disminuye, se necesita mayor detalle en los componentes del sistema.

Los modelos presentados en la Figura 2.3 corresponden a su versión más básica. A continuación se describe en términos generales cada uno de ellos. Esta descripción es importante porque entrega información de los elementos que se deben considerar para formular la metodología de selección que se presenta en el siguiente capítulo.

- Protección y estabilidad: a medida que la ventana de tiempo es menor, se deben detallar los distintos componentes del sistema para asegurar que la operación sea confiable y segura ante perturbaciones y fallas. En la escala de los milisegundos se analiza la estabilidad angular, frecuencia y voltaje. También se utiliza para diseñar los tiempos de operación de los distintos elementos de protección y sistemas de control [21].
- Flujos de potencia: es el modelo más importante y esencial para investigar problemas en la operación y planificación de los SE. Basándose en una condición de operación específica, es posible conocer el estado de la red (voltajes y flujos por las líneas). Se considera que la red se encuentra en régimen estacionario, por lo tanto las ecuaciones son algebraicas y no lineales [22].
- Despacho económico: en este problema se busca determinar cuál será la potencia que deberá ser suministrada por cada generador para abastecer la demandada. El criterio consiste en minimizar los costos de operación del sistema. En su versión más simple el sistema no considera la red de transmisión, pero sí algunas restricciones de las unidades generadoras. También es posible considerar el sistema de transmisión a través de modelos de flujo DC, lo que es usual en los modelos actuales [23].
- Predespacho: el objetivo es determinar qué unidades estarán encendidas en el siguiente día o semana. Este problema requiere considerar un gran número de restricciones técnicas de los generadores implicando un gran desafío en el problema de optimización. Esta dificultad combinada con el beneficio del ahorro en los costos operacionales, ha llevado este problema a ser analizado constantemente [24]. El predespacho es importante para el análisis de integración de ERV porque ha sido la herramienta más utilizada para evaluar el impacto de las ERV en los SE.
- Mantenimiento: los generadores térmicos típicamente requieren entre 1 a 5 semanas de mantenimiento por año y no se encuentran disponibles durante ese tiempo. El mantenimiento es programado en periodos de baja demanda eléctrica. Se debe garantizar que aún durante los periodos de baja demanda, exista capacidad disponible para operar en forma segura y confiable. Los modelos de optimización tienen que considerar restricciones técnicas y de seguridad [25]
- Coordinación hidrotérmica: se busca programar el uso de los embalses y el suministro de la demanda eléctrica. En mediano y largo plazo se resuelve un problema de optimización estocástico multietapas, el cual es complejo debido a la incertidumbre hidrológica. Su resolución ha sido fuente de extensa investigación. Dada su complejidad se realizan simplificaciones, la demanda se modela a través de curvas de duración y se desprecian los efectos de las restricciones de corto plazo de los generadores [26]
- Costos de producción: se analizan para horizontes de tiempo de un año o mayor, considerando restricciones de corto plazo como las del predespacho, y a ello se agregan limitaciones por mantenimiento de las unidades. En términos del modelo, es un problema de optimización complejo debido al tamaño, pero es un problema extendido de predespacho para sistemas predominantemente térmicos. En sistemas con centrales hidroeléctricas es utilizado el modelo de optimización de la coordinación hidrotérmica [23].
- Planificación de la expansión de SEP: en este problema se pretende optimizar las inversiones de tal forma que sean minimizados los costos de inversión y operación del sistema junto con asegurar el abastecimiento y la seguridad del SEP. Existen dos problemas relacionados de expansión: generación y transmisión, los que usualmente se han desarrollado en forma separada por la complejidad. Sin embargo, existen diversas propuestas que abordan este problema en conjunto [27].

Los problemas descritos en los ítems anteriores pueden ser reagrupados en problemas de planificación de la expansión, planificación de la operación y operación real como se observa en la Tabla 2.3. Desde una mirada matemática los problemas de planificación de expansión y operación se formulan a través de

modelos de optimización lineal, lineal entero-mixto y en sus versiones determinísticas, robustas y estocásticas. Para cada una de estas formulaciones las ERV corresponden a datos de entrada, los que deben ser compatibles con el horizonte que se está evaluando en el problema. Por ejemplo, para un predespacho day-ahead, si la ventana de tiempo es una hora, entonces las series de recurso a lo más deben tener un tiempo de resolución horaria.

En cada uno de los problemas de SE las ERV son datos de entrada en los modelos. También existen distintas suposiciones en la formulación, por ejemplo cuando se habla de problemas de mediano a muy largo plazo es usual considerar la demanda a través de la curva de duración, lo que implica un desacople temporal del problema y enfrentarlo con una perspectiva energética. No obstante, debido al aumento de las ERV en los SEP se reconoce la necesidad de incorporar simulación cronológica en problemas de largo plazo. Una aproximación para considerar las restricciones técnicas consiste en simular semanas o días tipo para el horizonte que se desea abordar. Para ambos casos se hace necesario una serie de los recursos cuya longitud sea al menos igual al horizonte que se desea resolver y esto es independiente del tipo de simulación que se esté realizando. La red de transmisión suele ser simplificada a un conjunto de líneas que representen los flujos principales.

En los problemas de planificación de la operación de corto y muy corto plazo las ERV siguen siendo datos de entrada, el detalle de los elementos de los SEP es mayor, pero los problemas matemáticos que son resueltos son predespacho y flujos de potencia. Las series de ERV que se utilizan como datos de entrada tienen una longitud menor, pero deben ser representativas de la variabilidad e incertidumbre del recurso. Como práctica se utilizan las series de pronósticos esperados, sin embargo cuando los niveles de ERV son elevados solo considerar el escenario esperado puede conducir a dificultades en la operación [12].

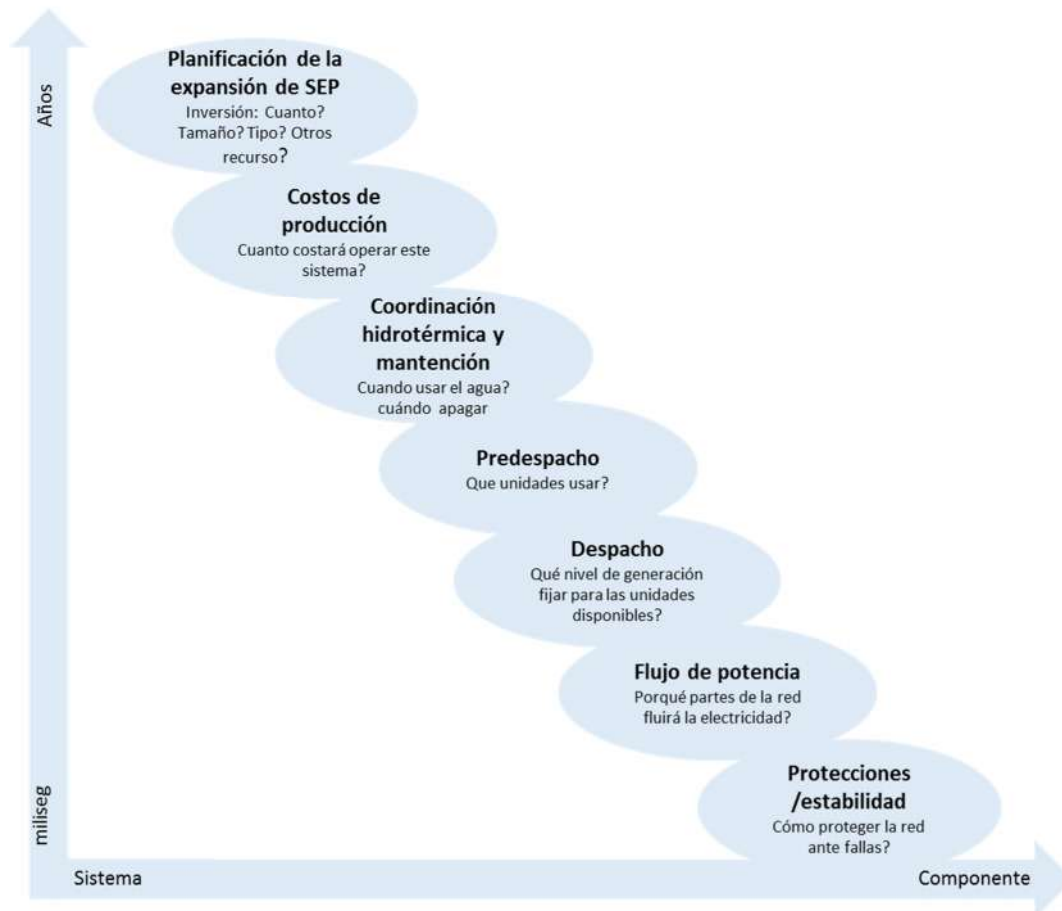


Figura 2.3. Modelos de planificación y operación en función del detalle y escala de tiempo considerados [28]

Tabla 2.3. Descripción de problemas de operación y planificación según distintos horizontes de evaluación [28]

Problema de SEP	Horizonte	Funciones
Planificación de la expansión de SEP	Muy largo plazo: de 5 hasta 25 años	Expansión de generación y transmisión Análisis de nuevas plantas y retiro de equipos
Planificación de la operación	Largo plazo: de 2 a 5 años	Establecer contratos Programación del manejo de embalse multianuales
	Mediano plazo: desde 1 mes hasta 2 años	Programar el mantenimiento Programación del manejo de embalses Producción de costos Modelos de confiabilidad
	Corto plazo: entre 1 a 4 semanas	Programación semanal del encendido y apagado de las unidades térmicas
	Muy corto plazo: menor a 1 semana	Pre despacho para todas las unidades generadoras
Operación en tiempo real	Menor a 1 hora	Despacho económico Control de frecuencia y voltaje

2.4 Impacto en la energía solar y eólica

La integración de energía eólica y fotovoltaica agrega variabilidad e incertidumbre al SE. El término variabilidad hace referencia a que la generación de electricidad en base a estas fuentes varía de acuerdo a la disponibilidad del recurso, la cual no es controlable y su almacenamiento aún presenta desafíos. La incertidumbre existe en el hecho que para cada instante de tiempo la magnitud exacta del recurso no se puede conocer en forma exacta [2].

En esta sección se comentan las características de la energía solar y eólica así como el impacto que tienen en los sistemas eléctricos. El enfoque se centra en aquellos aspectos que generar desafíos en la operación y planificación. No se comentan aspectos fenomenológicos de las ERV, aquellos aspectos pueden ser revisados con detalles en las referencias [29]–[31].

2.4.1 Características

La radiación solar y la velocidad de viento pueden tener variaciones a lo largo de segundos, minutos, horas, días, meses. Por ejemplo, la radiación solar experimenta cambio debido a la posición del sol a lo largo del día y de las estaciones del año. Estos fenómenos son bien conocidos y pueden ser modelados con precisión [30]. La salida y puesta del sol lleva a cambios en la generación que en promedio llegan a un 10% de la capacidad instalada en períodos de 15 minutos. La incertidumbre en la radiación es ocasionada por la nubosidad y los impactos en la variabilidad son observados con mayor detalle en el corto plazo: de los segundos a minutos. Los cambios debido a la nubosidad pueden exceder el 60% de la capacidad instalada [32].

La variabilidad de corto plazo ocasionada por la nubosidad puede ser reducida al aumentar la distancia entre los paneles, permitiendo que el impacto sea menor. En la Figura 2.4 se observan funciones de distribución acumulada de las variaciones ($\Delta = p(t) - p(t - k)$) para una planta de 30 kW y para múltiples plantas. De ella es posible observar que la probabilidad de ocurrencia de grandes cambios disminuye cuando se consideran múltiples plantas. A medida que se aumenta el tiempo, las variaciones son menores para las plantas de mayor tamaño, este efecto es debido al agregado de las plantas que están ubicadas en distintos sectores geográficos. Desde un punto de vista estadístico a medida que la distancia entre las plantas aumenta se ha observado que la correlación disminuye en forma exponencial como se observa en la Figura 2.5. Al disminuir la correlación entre las plantas se obtienen series con menor variabilidad [3].

La velocidad del viento y/o plantas eólicas tiene patrones más complejos; en algunos casos es posible definir variaciones diarias un poco más marcadas, pero es altamente dependiente de los lugares bajo estudio, diferente a los casos de la radiación, donde se tiene completa certeza del ciclo diario. En la Figura 2.6 se muestra un diagrama de cajas para series de velocidad de viento obtenidas a través del explorador eólico de

la Universidad de Chile. El punto se encuentra cerca de Calama en la segunda región de Antofagasta. Es posible observar que existen variaciones intradiarias en la media y varianza. Se debe tener presente que estas variaciones son completamente dependiente del lugar y no necesariamente pueden existir grandes variaciones estacionales para cada una de las series de velocidad.

La variabilidad de la generación eólica al igual que la generación fotovoltaica disminuye al considerar la agregación de diferentes plantas ubicadas en diversos lugares geográficos[3]. Los coeficientes de correlación en función de la distancia entre distintas plantas también se observa en las plantas eólicas y se ha informado en los trabajos [33][34]. En un estudio realizado por NREL se caracterizó el comportamiento de las fluctuaciones en la generación eólica usando registros de 35 centrales ubicadas en diferentes zonas de Estados Unidos [35]. De acuerdo a los datos, en escalas de 1 segundo los cambios son generalmente pequeños: 0.1% en promedio y en la mayoría de los casos son menores al 1%. En escalas de tiempo de 10 minutos la variabilidad se hace más significativa siendo en promedio 2% de la capacidad instalada y en el 98% de los casos los cambios en la generación son menores al 11%.

Para minimizar el efecto de la incertidumbre en la operación de SE para distintos horizontes de tiempo se han desarrollado modelos de predicción. Independiente de los grandes avances desarrollados en dichos modelos, siempre existe error en las predicciones y éste se ve incrementado a medida que el horizonte de pronóstico aumenta. Para el caso del viento se tienen errores entre 5% y 7% para un horizonte entre 1 y 2 horas. Los errores pueden llegar hasta el 20% para predicciones de un día [36]. Para el caso de la predicción de radiación, ésta suele ser más certera. Sin embargo, en algunas zonas geográficas podría no darse esta situación por la presencia de la nubosidad, implicando mayor variabilidad e incertidumbre, por lo tanto más complejidad en el proceso de predicción. Los errores de predicción tienen distintos impactos en la operación y depende del nivel de ERV presente en el SE [37]. El error de pronóstico en ERCOT ISO² en promedio es aproximadamente 10% de la capacidad instalada de energía eólica, la cual fue 9838 MW a principios del 2012. Este escenario de incremento en los niveles de ERV y errores típicos de predicción, están llevando a la necesidad de cambiar el paradigma de la operación y planificación de SE, pasando de un enfoque determinista a uno probabilístico. Las series sintéticas son utilizadas en la generación de escenarios para la toma de decisiones bajo incertidumbre [12].

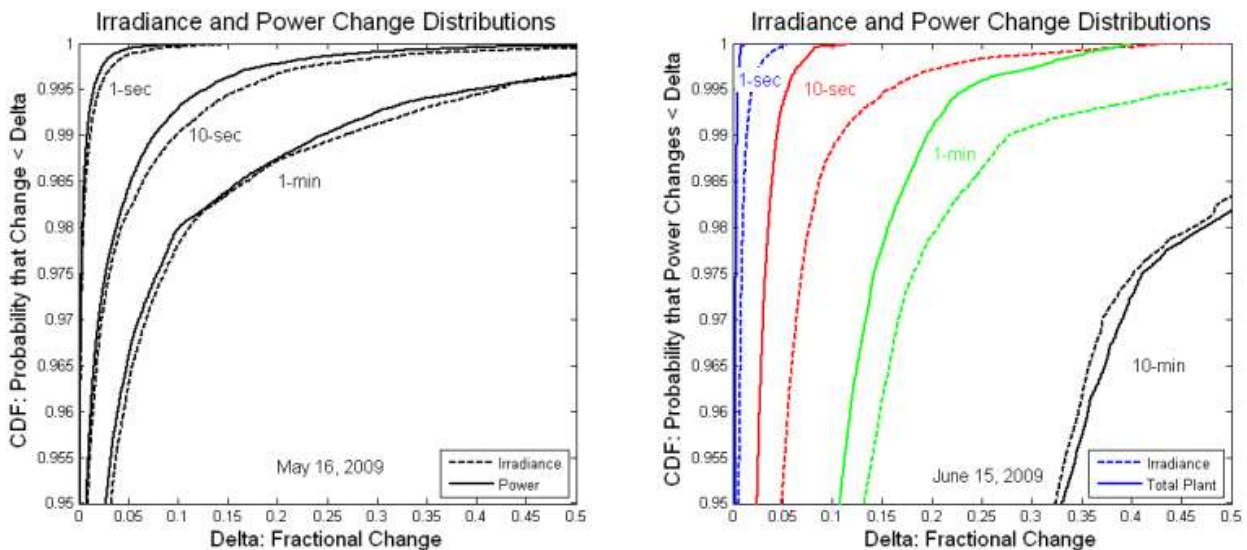


Figura 2.4. Función de distribución acumulada (entre los percentiles 95% y 100%) de cambios en radiación y potencia fotovoltaica para varios periodos de tiempo durante 1 día para un sistema de 30 kW fotovoltaicos (izquierda) y un sistema múltiple del orden de los MW fotovoltaicos (derecha) [32]³

² Es el operador de sistemas en Texas (del inglés *Electric Reliability Council of Texas*)

³ CDF función de distribución acumulada (del inglés *Cumulative Distribution Function*)

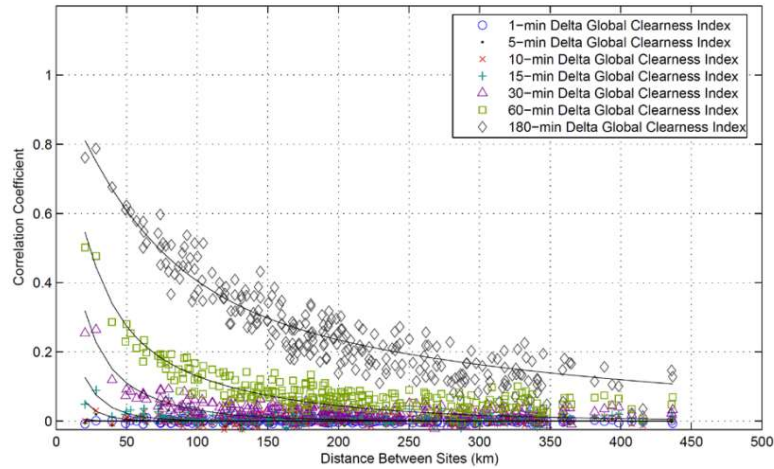


Figura 2.5. Coeficientes de correlación para cambios en el índice de claridad para diferentes distancias entre sitios y diferentes intervalos de tiempo [32].

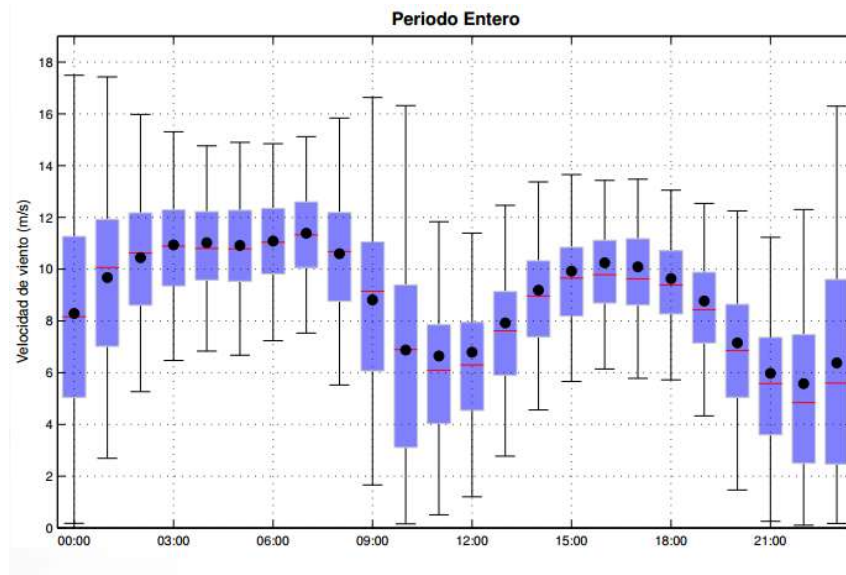


Figura 2.6. Ciclo diario de velocidad de viento a 90 metros. Los puntos negros indican el promedio de viento en cada hora del día. Las líneas rojas indican el valor mediano de las distribuciones horarias. Las barras azules representan el rango intercuartil. Las barras negras indican el rango de los valores horarios excluyendo valores extremos.

2.4.2 Impacto en la operación de sistemas eléctricos

El impacto en la operación debido a la integración a gran escala de las ERV debe ser analizado en diferentes intervalos de tiempo (segundo a segundo, minuto a minuto, intrahoraria, horaria, diaria, semanal y anual). En general los impactos en los SE son diferentes dependiendo del intervalo de tiempo que se observe. Ello implica que la integración de ERNC debe ser analizada e incluida en los procesos de planificación de corto y largo plazo [38]–[40].

La correlación entre las ERV depende de la geografía en donde se encuentran instaladas las centrales y ocasiona que la demanda neta del sistema se vea modificada agregando mayor variabilidad. Debido a que la curva de la demanda neta es diferente, la planificación de la operación para los generadores convencionales cambia. En la Figura 2.7 se muestra una gráfica de ejemplo, en donde es posible observar

el aumento en la variabilidad en la curva de demanda neta a causa de la integración de ERV. Tres efectos se pueden observar en la figura: aumento de la incertidumbre ocasionada por los pronósticos, aumento en la magnitud de las rampas de potencia y aumento de las frecuencias de dichas rampas.

En los SE cuya matriz energética es principalmente en base a centrales térmicas y sin sistemas de almacenamiento a gran escala, la ERV ocasionan una modificación significativa en el predespacho de las unidades generadoras: se disminuyen costos de operación debido a que el costo de operación de las ERV es cero, pero por otro lado, aumenta el encendido y apagado de las unidades térmicas. Para aquellas situaciones en donde la generación de ERV sea capaz de suministrar toda la demanda, las centrales térmicas deberían apagarse u operar a mínimo técnico. Si los generadores térmicos se desconectan no podrían estar presentes para abastecer la demanda cuando no exista disponibilidad de las ERV, debido a límites técnicos de tiempos mínimos de encendido y apagado, especialmente para las centrales a carbón. Estas situaciones pueden ocasionar que el SE recorte la generación variable y con ello disminuya la participación de las ERV [38][41]

Como se ha mencionado, con la variabilidad de las ERV el número de encendido y apagado de las unidades térmicas se ve incrementado. En la Figura 2.8 es posible observar que las unidades de ciclo combinado se ven más afectadas debido a la integración de las ERV, aumentando en gran medida los encendidos. En el caso de las unidades de carbón, el número de encendidos aumenta levemente, pero decae cuando los niveles de ERV son elevados, ello ocurre porque es necesario mayor reserva en giro para proveer la flexibilidad que necesita el SE para operar en forma segura [42]. El aumento de encendidos en las centrales de ciclo combinado es debido a que son unidades que poseen mayor flexibilidad y tiempos mínimos de apagado y encendido menores a las centrales de carbón.

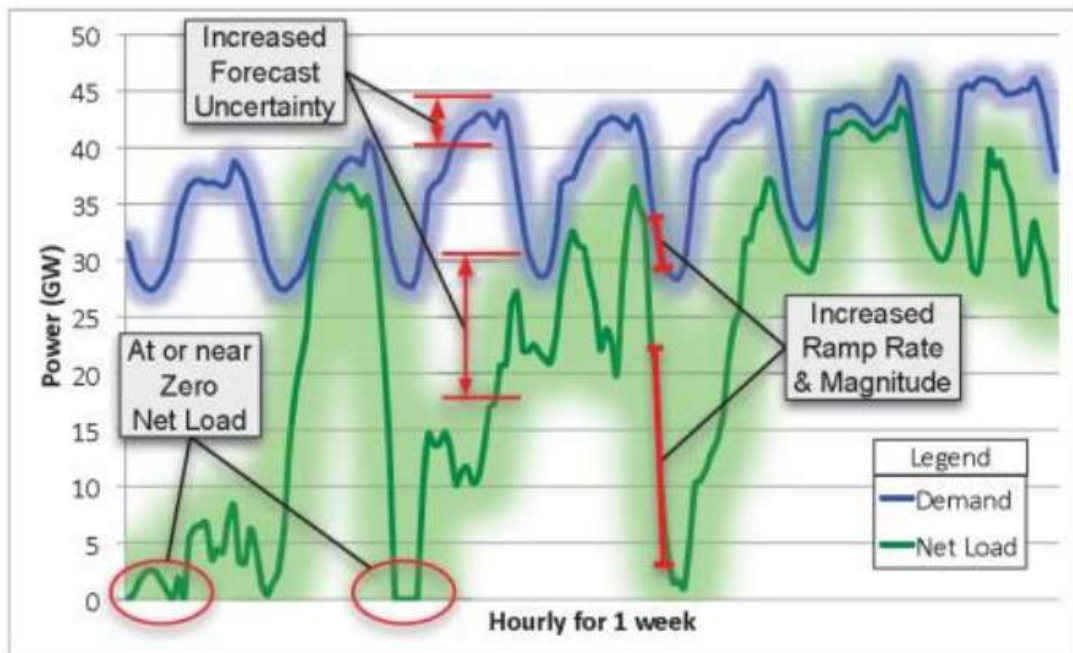


Figura 2.7. Gráfica de la demanda del sistema (azul) y la demanda neta (verde) en donde se ejemplifica distintos impactos en los SE debido a la integración de ERV[28].

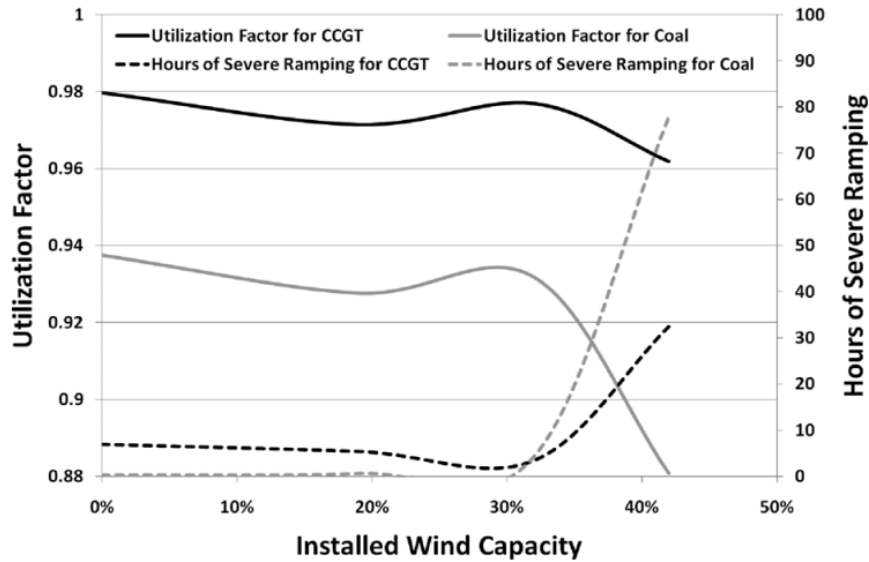


Figura 2.8 Número de horas donde las unidades de carbón y de ciclo combinado aumentaron o disminuyeron severamente su generación [42]

2.4.3 Impactos en la planificación de sistemas eléctricos

En los problemas de planificación de SE se busca definir la inversión de generación, el sistema de transmisión o ambos. Éste es un problema de optimización complejo lo que ha llevado a realizar una serie de simplificaciones. El enfoque tradicional considera la demanda a través de una curva de duración [43], lo que conlleva a que sean despreciadas las restricciones técnicas de tiempos de partida y encendido de las unidades. Para sistemas completamente térmicos esta suposición no afecta considerablemente los resultados porque la demanda experimenta variaciones lentas. Cuando los niveles de penetración de ERV son a gran escala, esta representación podría no ser adecuada. Como fue descrito en la sección anterior a mayor variabilidad mayores los cambios en los encendidos y apagados de las unidades, lo que implica que el SE necesita mayor flexibilidad. Así, las nuevas unidades deben cumplir con las siguientes características[44]:

- Amplio rango entre su potencia mínima y máxima de operación.
- Alta capacidad de toma y bajada de carga.
- Tiempos mínimos de operación reducidos.

El problema de planificación considerando flexibilidad para todas las unidades, puede llevar a resultados en donde centrales a carbón que deban ser instaladas no dispongan de la flexibilidad necesaria para lidiar con la variabilidad y se tenga que recortar ERV. Estas unidades, una vez encendidas, deben mantenerse operando por varios días y una vez apagadas, no pueden encenderse de forma inmediata. Tampoco pueden tomar carga rápidamente [44]. Para entender este efecto se han realizado algunos estudios que han tenido como objetivo evaluar el impacto que tiene considerar los efectos de corto plazo en un ejercicio de planificación. En [45] se generan manualmente diferentes planes de inversión y se calcula su costo total, costos de operación y costo de inversión, utilizando un modelo de lista de mérito y un modelo de preespacho para representar la operación para distintos niveles de generación variable. Los resultados muestran que cuando la penetración de ERV es baja, los planes de inversión que minimizan el costo de operación determinados por ambos modelos son similares. Sin embargo, en presencia de altos volúmenes de ERV, las diferencias son significativas, ya que el modelo de lista de mérito no logra identificar la necesidad de flexibilidad e instala mayor cantidad de unidades inflexibles. En [4] se desarrolla una herramienta computacional para considerar las ERV y se aplica un caso para el SING. En los resultados se muestra que los costos totales del sistema disminuyen porque se toma la decisión de invertir menos en ERV para diseñar un sistema con mayor flexibilidad, además los recortes de ERV también decrecen.

2.5 Generación de series sintéticas

En esta sección se realiza una revisión bibliográfica respecto a los distintos modelos o metodologías que han sido desarrollados en la literatura para generar series sintéticas de velocidad de viento, potencia eólica, radiación solar y potencia fotovoltaica. La separación se hace en función del tipo de dependencia y recurso que se ha caracterizado en el modelo.

Una serie sintética es una serie de tiempo generada a través de una ecuación matemática normalmente de origen empírico, que representa un proceso estocástico. Estas deben ser capaces de replicar los estadísticos de la serie real tales como: la media, varianza, función de densidad de probabilidad, dependencia temporal y dependencia espacial [46]. Cada recurso es considerado como un proceso estocástico. En palabras sencillas un proceso estocástico se puede definir como un conjunto de variables aleatorias $\{X_t, \text{ con } t \in T\}$, ordenadas según el subíndice t [47] (para mayor detalle de dependencia y procesos estocásticos ver Anexo 1).

Para cada uno de los recursos que aborda este trabajo de tesis se consideran las siguientes propiedades estadísticas:

- Poseen variaciones estacionales
- La función de densidad de probabilidad no es Normal
- No son series estacionarias
- Tienen correlación temporal y espacial
- La transformación de la velocidad del viento a potencia eólica es una función no lineal

La revisión empieza con los modelos desarrollados para generar series sintéticas que buscan caracterizar solo la dependencia temporal y luego dependencia espacio temporal. La forma en la cual se ha buscado en la literatura medir la dependencia temporal es a través de las funciones de autocorrelación simple y parcial, y la matriz de autocorrelación cruzada para medir dependencia espacial. Estos conceptos son presentados con detalle en el anexo 1 de esta tesis, los cuales deberán ser revisados para entender adecuadamente el contenido de esta sección en caso de estar familiarizados. De igual forma se recomienda el estudio de dicho anexo para repasar conceptos de estacionariedad, proceso estocástico, entre otros.

Cuando se hable de modelos en la revisión, se deberá entender tanto la estructura matemática como los preprocesamientos y transformaciones que se hayan propuesto en la literatura. Como será expuesto durante esta sección, en diferentes estudios se han utilizado las mismas estructuras matemáticas pero variando tanto el preprocesamiento y transformaciones.

El objetivo de esta revisión consiste en poder conocer los requisitos que deben ser cumplidos por las series de tiempo de ERV para utilizar los modelos existentes y observar las ventajas y desventajas al ser aplicados a diferentes recursos.

2.5.1 Modelos para generar series sintéticas que consideran dependencia temporal

Cadenas de Markov

Las cadenas de Markov homogéneas son utilizadas con gran frecuencia para generar series sintéticas de velocidad de viento y potencia eólica. Una cadena de Markov es un proceso estocástico que cumple con la propiedad markoviana según se muestra en las ecuaciones (2.1) y (2.2), en donde se expresa que la probabilidad condicional del estado futuro sólo depende del estado actual. El modelo queda definido cuando se estimada la matriz de transición de probabilidades de la ecuación (2.3)

$$P\{X(t_n) = i_n | X(t_1 = i_1), \dots, X(t_{n-1} = i_{n-1})\} = P\{X(t_n) = i_n | X(t_{n-1}) = i_{n-1}\} \quad (2.1)$$

$$P\{X(t) = j | X(s) = i\} = P_{ij}(s, t) \quad (2.2)$$

$$P_{t,t+1} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,k} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k,1} & p_{k,2} & \cdots & p_{k,k} \end{bmatrix} \quad (2.3)$$

Cuando se habla de una cadena homogénea se refiere a que la matriz de transición no depende del tiempo. Las cadenas de Markov homogéneas son utilizadas cuando no existen variaciones estacionales muy pronunciadas en las series [48]. En [48] se presenta el algoritmo y se aplica a series de velocidad de viento en diferentes regiones de Turquía. Se utiliza una cadena de primer orden y entrega muy buenos resultados en la reproducción de la media y la desviación estándar. En [49] se aplica de nuevo el algoritmo en Tangier, Marruecos. A diferencia del trabajo anterior se comparan los histogramas y la función de autocorrelación simple mostrando que la reproducción de la función de densidad, la cual se representa a través del histograma, es similar a la serie real. Un avance se presenta en [50] donde se aplica una cadena de primer y segundo orden a dos sitios de Malasia manteniendo los buenos resultados mostrados en los trabajos anteriores y concluyendo que al aumentar el orden mejora la reproducción de la función de autocorrelación simple, sin embargo queda de manifiesto que el aumento de orden incluye un crecimiento exponencial de la matriz de transición de probabilidades lo que incrementa la complejidad del modelo.

Las cadenas de Markov necesitan un espacio de estados y éste fue definido utilizando distintos criterios. En [51] se analiza el impacto que tiene la cantidad de estados al sintetizar la serie. Se aplica el modelo a series de velocidad de viento para variaciones de 1 m/s y 0,5 m/s mostrando que existe una mejora en la reproducción de los estadísticos. No se comenta de qué forma esto impacta la función de autocorrelación simple y parcial.

Es usual ajustar los modelos de series sintéticas a series de velocidad de viento y luego transformarlas a series de potencia. Ésta transformación es bastante compleja debido a la no linealidad de la curva velocidad potencia de las turbinas, a ello se agrega el efecto parque, el cual consiste en la disminución de la variabilidad total del parque debido a la correlación que existe entre las distintas turbinas en función de la distancia, como fue observado en la sección anterior. Al ver estas dificultades se propuso utilizar cadenas de Markov directamente a series de potencia para resoluciones de tiempo: 15, 30 y 60 minutos [52]–[54]. Se concluye que la cantidad de parámetros disminuye debido a que la función de densidad de una turbina eólica tiene una concentración de observaciones para las cuales la potencia de salida es cero, como se muestra en la Figura 2.9. Se observa que las cadenas de Markov pueden generar series sintéticas con funciones de densidad no Gaussianas [53][55].

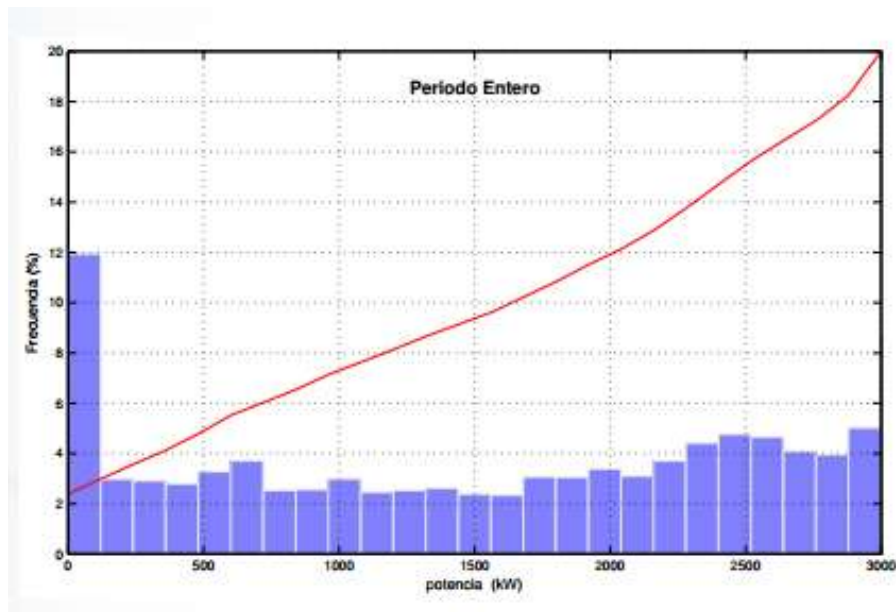


Figura 2.9. Distribución de frecuencia de la potencia del aerogenerador Vestas V126-3.0 MW a 90 metros. Las barras azules representan el porcentaje de valores dentro de cada intervalo. La línea roja es la distribución de frecuencia acumulada⁴

Dentro de las principales falencias de las cadenas de Markov se encuentra que no modelan adecuadamente la función de autocorrelación simple debido a que se asume que el estado futuro sólo depende del estado actual para el caso de primer orden o del actual y el anterior para el caso de segundo orden. En [54] se hace una advertencia a la utilización de este modelo para generar series sintéticas. En él se expone que si el tiempo de resolución es menor a 30 minutos el error en la función de autocorrelación simple aumenta considerablemente.

En los trabajos [56]–[58] se proponen mejoras al modelo, las cuales permiten superar algunas de las deficiencias descritas en los párrafos anteriores. Se propone utilizar una cadena por estación para considerar variaciones estacionales marcadas, lo que implicaría utilizar un modelo no homogéneo pero con mayor cantidad de parámetros. También se propone una metodología para considerar series con tiempo de muestreo menor a una hora.

En el caso del recurso solar, existen menos trabajos desarrollados que utilizan cadenas de Markov. Debido a que las series de radiación o potencia fotovoltaica contienen fuertes componentes estacionales no es posible modelar la radiación o potencia fotovoltaica con una cadena de Markov homogénea, por lo tanto la estrategia consiste en modelar a través de cadenas de transición por estación. En [20] y [21] se aplica el modelo a series de índice de claridad k_t , el cual se define como:

$$k_t = \frac{H}{H_e} \quad (2.4)$$

Donde H es radiación global horizontal y H_e es radiación extraterrestre. En ambos trabajo se obtienen buenos resultados en replicar las propiedades estadísticas pero se reconoce como desventaja la gran cantidad de parámetros que necesita el modelo.

En resumen, este modelo puede representar diferentes funciones de densidad así como los estadísticos. Sin embargo, se debe tener cuidado en la caracterización de la función de autocorrelación simple. Es normal

⁴ El gráfico fue obtenido utilizando el explorador eólico en una zona cercana a la ciudad de Calama en la II región de Chile

encontrar en la literatura que este modelo no representa adecuadamente esta función, sin embargo, esa declaración es relativa: depende del comportamiento específico de la serie que se esté estudiando. Si la serie bajo estudio no tiene variaciones intradiarias muy marcadas la función de autocorrelación será modelada adecuadamente. Para el caso en que sí existan correlaciones durante el día habrá un mayor error en la caracterización de la función de autocorrelación. Las Figura 2.10 y 2.11 muestran comparaciones entre las funciones de autocorrelación simple reportadas en dos trabajos. En la Figura 2.10 es posible apreciar que la serie de velocidad de viento posee correlaciones marcadas para los retardos múltiplos de 24, lo que indicaría que la serie experimenta una variación intradiaria. Para este caso las cadenas de Markov sólo pueden replicar la persistencia para los primeros 5 retardos. En cambio, en la Figura 2.11 la serie de energía eólica no presenta estos peak para el retardo 24, más bien, se observa un decaimiento exponencial, el cual puede ser modelado con menos error a través de las Cadenas de Markov. Estos dos ejemplos verifican lo expuesto anteriormente.

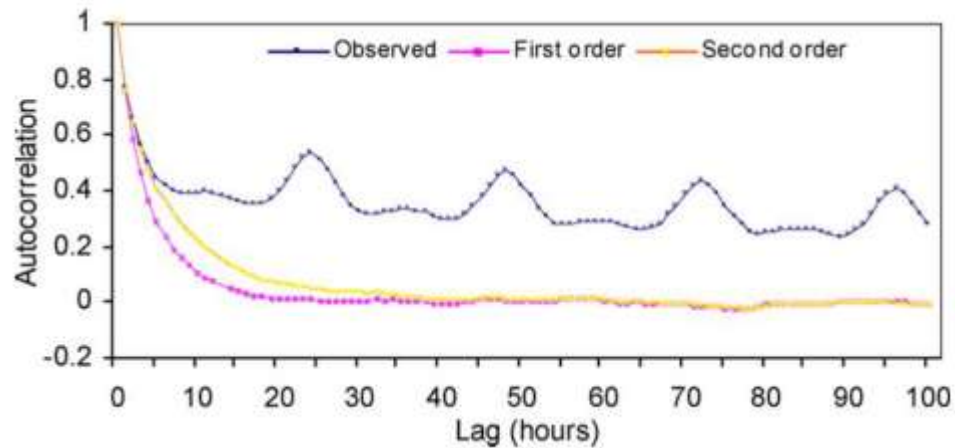


Figura 2.10. Comparación de las funciones de autocorrelación simple entre la serie real y la serie sintética generada utilizando cadenas de Markov de 1° y 2° orden reportado en [50]

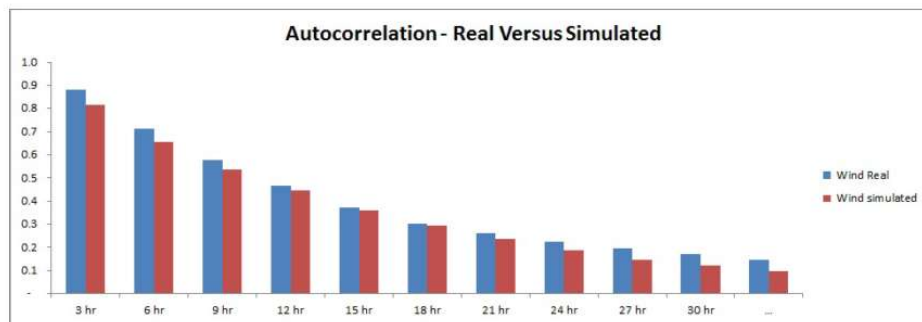


Figura 2.11. Comparación de las funciones de autocorrelación simple entre la serie real y la serie sintética generada utilizando cadenas de Markov de 1° reporado en [61]

Modelos ARMA⁵ (p,q) y ARIMA(p,d,q)⁶

Los modelos ARMA (p,q) y ARIMA(p,d,q) han sido desarrollados durante las últimas décadas siendo de gran utilidad para explicar diferentes fenómenos en econometría, series climatológicas entre otras. Los modelos ARMA (p,q) pueden ser descritos mediante la ecuación (2.5), donde los parámetros ϕ_i, θ_j con $i = 1, \dots, p$ y $j = 1, \dots, q$ corresponden a los coeficientes auto regresivos y de media móvil del modelo. ε_t

⁵ Del inglés *AutoRegressive Moving Average*

⁶ Del inglés *AutoRegressive Integrated Moving Average*

corresponde a un proceso ruido blanco que es independiente e idénticamente distribuido de acuerdo a una distribución normal $N(0, \sigma_\varepsilon)$

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (2.5)$$

Para poder aplicar correctamente este modelo la serie de tiempo debe ser estacionaria (ver Anexo 1) de lo contrario es necesario aplicar transformaciones para cumplir con esta condición. Una de las transformaciones que se realiza para convertir una serie de tiempo que posee tendencia consiste en diferenciar la serie, de esta forma el proceso descrito en la ecuación (2.5) puede ser descrito en forma general a través de las siguientes expresiones:

$$X_{t-i} = B^i X_t \quad (2.6)$$

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d X_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (2.7)$$

Donde B^i es denominado el operador retardo y actúa sobre la serie de tiempo. En caso que exista tendencia, sería posible tener una nueva serie de tiempo sin presencia de tendencia y de esta forma obtener un proceso estacionario. La obtención del orden p y q es a través de criterios de información, en especial el criterio AIC (AIC por Akaike's Information Criterion).

En la literatura existen numerosos trabajos en los cuales se han aplicado estos modelos para generar series sintéticas de velocidad de viento, potencia eólica y radiación [8], [10], [12], [30], [33], [34], [43], [46], [62]–[82]. La estructura matemática es idéntica pero cada modelo difiere en el pre procesamiento de los datos para lograr quitar estacionalidades, tendencias y transformar las series a estacionarias. Generalmente se han aplicado en series cuyo tiempo de muestreo va desde 1 hasta 24 horas [30].

En los modelos se propone que la serie de tiempo del recurso es posible descomponerla en una componente determinista más una componente estocástica:

$$X_t = T_t + S_t + R_t \quad (2.8)$$

T_t , S_t corresponden a la tendencia y estacionalidad respectivamente, las cuales son consideradas determinísticas. R_t corresponde a un proceso ARMA(p, q). A continuación se resumen las distintas propuestas para transformar una serie no estacionaria a estacionaria. Se debe tener presente que todas las propuestas son la aplicación de la teoría de series de tiempo desarrollada en los años setenta [83] y la innovación principalmente radica en la experimentación de los modelos a series de tiempo en sitios específicos.

Uno de los primeros trabajos propuestos corresponde al desarrollado en [79], en donde se propone estabilizar la varianza utilizando la transformación de Box-Cox [83] para luego quitar estacionalidades utilizando la expresión:

$$Y_t = \frac{X_t - u_{tk}}{\sigma_{tk}}, \quad \forall k \in \{S\} \quad (2.9)$$

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

En la ecuación (2.9) el conjunto S corresponde a la estacionalidad de la serie de tiempo, la cual podría ser, para el caso de series de tiempo horarias: $S = \{1, \dots, 24\}$ para variaciones estacionales horarias o $S = \{1, \dots, 12\}$ para variaciones estacionales mensuales.

El hecho de aplicar la transformación Box-Cox para estabilizar la varianza no garantiza que el nuevo proceso tenga distribuciones marginales normales. Este requisito es importante debido a que, si esta condición no se cumple, el modelo no será capaz de representar la función de densidad de probabilidad de la serie real. Esta metodología también es aplicada a series de velocidad de viento en [10], [33], [74] y [9]. Cuando las condiciones de normalidad se cumplen, el modelo entrega buenos resultados porque caracteriza adecuadamente la función de densidad de probabilidad y debido a la complejidad del mismo, la función de autocorrelación es mejor caracterizada comparado con los modelos de cadenas de Markov.

En [34] se presenta un modelo ARMA(p,q) aplicado a Inglaterra para estudios de integración de ERV en su SE. Tiene como objetivo modelar adecuadamente las variaciones anuales y diarias. Para ello definen una nueva serie:

$$Y_t = X_t - X_{at} - X_{dt} \quad (2.10)$$

Donde X_t , X_{at} , X_{dt} corresponden a la serie de tiempo real, variación anual y diaria respectivamente. Estas series son obtenidas a través de un ajuste de curva utilizando series de Fourier para luego ajustar un modelo ARMA (p,q) para el caso univariable y se extiende para el caso multivariable (la versión multivariable será descrita en la siguiente sección). Se asume que al realizar esta resta el nuevo proceso se distribuye con una normal. No se aplican test estadísticos para verificar esta condición pero la gráfica de los histogramas lucen aproximadamente a una normal [34]. El modelo es aplicado a series de velocidad de viento con tiempo de muestreo de 1 hora.

En [84] se presenta una metodología para caracterizar la dependencia espacio temporal. Se transforma la serie de velocidad de viento no estacionaria utilizando la transformada integral. Dicha aproximación consiste en aplicar la función de distribución de probabilidad acumulada, con ello se genera una nueva variable aleatoria que se distribuye en forma uniforme para luego aplicar la función normal inversa. Esta aproximación es independiente del proceso para caracterizar la relación espacial, por lo tanto es aplicable solo a modelos univariables. Una extensión a dicha metodología es realizada en [9] en donde agrega un procesamiento previo para remover la estacionalidad utilizando la ecuación (2.9). El proceso queda descrito mediante el siguiente conjunto de ecuaciones:

$$\begin{aligned} X'_t &= \frac{X_t - u_{tk}}{\sigma_{tk}}, \quad \forall k \in \{S\} \\ Y_t &= F_{invN}^{-1} \left(F \left(X'_t \right) \right) \\ Y_t &= \sum_{i=1}^p \phi_i Y_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \end{aligned} \quad (2.11)$$

Donde F_{invN}^{-1} es la función de distribución acumulada de probabilidad normal estándar, F es la función de distribución acumulada de probabilidad de la serie de tiempo X'_t y X'_t es la serie sin la componente estacional y por último Y_t es la serie transformada a la cual se ajusta el modelo.

También se ha propuesto utilizar una función de distribución por estación para transformar la serie a estacionaria, tal como se presenta a continuación [85][78][86]:

$$\begin{aligned} Y_t &= F_{invN,k}^{-1} \left(F \left(X_t \right) \right), \quad \forall k \in \{S\} \\ Y_t &= \sum_{i=1}^p \phi_i Y_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \end{aligned} \quad (2.12)$$

La transformada integral ha sido utilizada en series de velocidad de viento, potencia eólica y potencia fotovoltaica [86][78]. Con esto, es posible garantizar que las series sintéticas serán capaces de replicar adecuadamente las funciones de densidad independiente de los recursos que se estén considerando. Anterior

a ello constantemente se argumentaba que los modelos ARMA(p,q) no eran capaces de caracterizar adecuadamente la función de densidad de probabilidades al ser comparado con las cadenas de Markov [77][84]. Pero con esta propuesta se supera esta desventaja. Sin embargo, surge otra y es que se necesitan suficientes datos para estimar adecuadamente la función de distribución de la serie. En el Anexo 1 se muestra gráficamente el funcionamiento de la transformada integral y su aplicación a las series de tiempo.

En [68] se propone utilizar un modelo ARIMA(p,d,q) para generar series sintéticas se series de potencia eólica. Se sigue el procedimiento de Box-Jenkins para determinar el orden del modelo. Una vez ajustado se agrega un limitador, acotando los valores máximos generados del modelo a la capacidad instalada de potencia eólica, de esta manera, se generan valores factibles de potencia eólica. El modelo es aplicado a series con resolución horaria, mostrando que se caracteriza adecuadamente la función de autocorrelación y estadísticos. No existe mayor información respecto a la consideración de la estacionalidad de la serie, sin embargo el autor en [8] propone las técnicas descritas en las ecuaciones (2.9) y (2.10) para remover estacionalidades.

De acuerdo a la revisión bibliográfica realizada en este trabajo, en este momento se desconoce si existe algún análisis o comparación entre las distintas propuestas. Las desventajas que se expusieron al iniciar la utilización de estos modelos se han superado a través de la transformada integral. La aplicación de la transformada Box-Cox no garantiza que el proceso nuevo sea normal después de su aplicación. Sin embargo, en los distintos trabajos de investigación, al aplicar esta transformación las funciones de densidad de probabilidad de la serie transformada se aproximaron a una normal. Una ventaja importante en estos modelos es que son más conocido y la estimación de sus parámetros no es compleja dado las capacidades de cómputo actuales.

El modelo ARMA(p,q)-GARCH(a,b) también es utilizado en economía para caracterizar series de tiempo que son altamente volátiles y han sido aplicados a series de velocidad de viento y radiación [85] [87] [88]. A diferencia de los modelos ARMA(p,q) existen menos trabajos en donde se use esta estructura matemática para generar series sintéticas. La aplicación de estos modelos se basa un análisis de residuos, en los cuales se observa que la varianza condicional no se mantiene constante. Al considerar esta característica, es posible analizar de mejor forma eventos extremos en las series de velocidad de viento u otro recursos que se esté analizando [85] . La estructura matemática de este modelo queda descrita a través de las siguientes ecuaciones:

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (2.13)$$

$$\varepsilon_t = z_t \sigma_t, z_t \sim N(0,1) \quad (2.14)$$

$$\sigma_t^2 = K + \sum_{j=1}^a \alpha_j \sigma_{t-j}^2 + \sum_{j=1}^a \beta_j \varepsilon_{t-j}^2 \quad (2.15)$$

Para superar los problemas de no normalidad, se utiliza la transformada integral. La estacionalidad horaria es removida al considerar una función de distribución de probabilidades por unidad de tiempo. Los resultados muestran que se reproducen adecuadamente los estadísticos globales tales como: función de densidad de probabilidades, función de autocorrelación.

Teoría de Cópula

La teoría de Cópula se utiliza para modelar la dependencia entre variables aleatorias. En los últimos años se ha utilizado para modelar dependencia espacial entre distintas centrales eólicas [89]–[93]. La teoría de Cópula relaciona la función de distribución conjunta con las funciones de distribución marginal de las variables aleatorias a través de una función (denominada Cópula). Esto puede ser expresado matemáticamente:

$$C[F_1(X_1), \dots, F_n(X_n)] = F(X_1, \dots, X_n) \quad (2.16)$$

Sea F y F_i , con $i = 1, \dots, n$ la función de densidad conjunta y marginal respectivamente. Entonces la función de densidad de Cópula c es definida por:

$$c(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial F_1(x_1) \dots \partial F_n(x_n)} \quad (2.17)$$

Consecuentemente, f puede ser expresada por:

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i) \quad (2.18)$$

Diferentes familias de Cópulas presentan variadas estructuras de dependencias, Cópulas elípticas tales como la Gaussiana y t son de fácil generalización en caso de considerar una gran cantidad de variables aleatorias [94]. En [95] se aplica por primera vez a una serie temporal de velocidad de viento para un tiempo de muestreo de 10 minutos. Se consideran 144 variables aleatorias para modelar la variabilidad horaria y estacional. Los resultados son comparados con un modelo ARMA(p,q) y con cadena de Markov. En este trabajo se observa que la teoría de Cópula caracteriza mejor estructuras de dependencia no lineal. A pesar de ello, se reconoce que los modelos son más complejos en la estimación de sus parámetros y el proceso de simulación.

Otros

Otros trabajos han aplicado redes neuronales perceptrón multicapas para generar series de radiación global horizontal en [96]. El tiempo de muestreo es de 1 hora y la variable aleatoria considerada es el índice de claridad. Se utiliza la capa de entrada para modelar la tendencia y una capa oculta para considerar la componente estocástica. El modelo es aplicado en siete sitios con climas muy variados. La evaluación del modelo se realiza a través de una comparación de las medias mensuales más el cálculo de la varianza relativa media, no existiendo comparación de funciones de autocorrelación ni comentarios respecto a las bondades del modelo.

En [63] y [64] se presenta una metodología para generar series de velocidad de viento la cual es desarrollada para generar series sintéticas a partir de estadísticos mensuales o anuales. El modelo propone una serie de ecuaciones para caracterizar variaciones estacionales diarias, anuales y componentes aleatorias. Para encontrar los parámetros del modelo se necesita resolver un problema de optimización multiobjetivo, para ello se utilizan algoritmos genéticos. Se realiza una comparación con cadenas de Markov, encontrando que tiene un mejor desempeño en la representación de la correlación temporal. La serie que estudian tiene un comportamiento similar al mostrado en la Figura 2.10 y dada las ecuaciones del modelo siempre se generaría una serie de tiempo cuya estructura de correlación tome esa forma.

Por último también se han utilizado métodos de análisis espectral en [99] y [77] entregando muy buenos resultados en modelar aquellas series de velocidades de viento con frecuencias de muestreo menor a una hora debido a que son capaces de representar de mejor forma las turbulencias.

2.5.2 Modelos para generar series sintéticas que consideran dependencia temporal y espacial

Cuando existen diversas centrales renovables esparcidas en sectores geográficos se ha reconocido la necesidad de considerar dependencia espacial y temporal en los modelos de generación de series sintéticas. Se han considerado dos tipos de modelos que caracterizan ambas dependencias. En primer lugar, los modelos multivariados autoregresivos VAR⁷⁷(p) [7], los cuales han sido aplicados bajo distintos supuestos en la literatura mostrando tener buenos resultados en la representación de los recursos renovables. Un segundo tipo de modelos corresponde a los que utilizan la teoría de Cópula, en donde queda de manifiesto

⁷⁷ Del inglés *Vector Autoregressive*

la complejidad matemática de su aplicación pero destacan las ventajas al modelar estructuras de dependencia más complejas como fue descrito en la sección anterior.

VAR(p)

Un modelo tipo VAR(p) tiene la siguiente estructura:

$$\vec{Y}_t = \sum_{i=1}^p A_i \vec{Y}_{t-i} + \vec{\varepsilon}_t \quad (2.19)$$

Donde A_i corresponde a la matriz de coeficientes modelo VAR (p), de orden $k \times k$, siendo k la cantidad de series de tiempo, y $\vec{\varepsilon}_t$ es un vector ruido blanco, el cual se distribuye en forma normal con matriz de covarianza no singular Σ . Para aplicar este modelo se debe cumplir que cada serie de recurso sea un proceso estacionario y se distribuya con una distribución normal. Esto último para caracterizar adecuadamente la función de densidad de probabilidad.

El modelo fue aplicado por primera vez a series de velocidad de viento y radiación en [86], [102]. Para poder considerar series de tiempo cuyas funciones de densidad fueran no normales y distintas se propone utilizar la transformada integral. Esta transformada es aplicada por hora con el propósito de quitar la variación horaria muy marcada del recurso solar. Se observa en los resultados que es posible caracterizar adecuadamente los estadísticos así como la dependencia entre los distintos sitios. Se reconoce que existe una complejidad en la cantidad de parámetros del modelo, la cual aumenta a medida que la cantidad de sitios aumenta.

Los modelos multivariantes también han sido aplicados a series de velocidad de viento en Inglaterra para estudios de integración de ERV. En [34] se aplica un modelo ARMA (p,q) y VAR(p) a distintos sitios de Inglaterra. Se remueven las estacionalidades de la serie a través de la sustracción de las medias anuales y horarias utilizando la ecuación (2.10). Este pre procesamiento permite que la nueva serie se distribuya en forma Gaussiana. El pre procesamiento aplicado a estas series para quitar la estacionalidad, no garantiza que las series se transformen a un proceso Gaussiano, sino más bien fue un hecho específico dada las características de las series utilizadas en dicho trabajo.

En [104], [7], [66] y [71] se aplica un modelo multivariable a series de potencia eólica en distintos países de Europa (Finlandia, Suecia y Dinamarca). En dichas series se observa una tendencia debido al aumento en la capacidad instalada de energía eólica. Para poder quitar dicha tendencia se define una nueva variable aleatoria en donde divide la serie de potencia por la capacidad instalada de cada país. La serie normalizada se convierte en un proceso Gaussiano a través de la transformada integral y la estacionalidad que poseen las series se modelan considerando una variable que depende del tiempo como se muestra en la siguiente ecuación:

$$\vec{Y}_t = \vec{c}_t + \sum_{i=1}^p A_i \vec{Y}_{t-i} + \vec{\varepsilon}_t \quad (2.20)$$

Un aporte en este trabajo radica en la forma de estimar las funciones de distribución al utilizar la transformada integral y en la modelación de los residuos. Se aplica la distribución Pareto generalizada a las colas superiores e inferiores de la función de distribución marginal. También se utilizan la familia de Cópulas t para muestrear el ruido blanco $\vec{\varepsilon}_t$. Los resultados muestran que la función de autocorrelación cruzada resulta idéntica a la real, de igual forma a los demás estadísticos.

CARMA(p,q)⁸

Reconociendo que la complejidad del modelo es creciente a medida que la cantidad de sitios aumenta, se ha propuesto una familia de modelos denominados en la literatura CARMA(p,q). Estos modelos son ampliamente utilizados en series de tiempo de afluentes y son una simplificación de los modelos VAR(p)

⁸ Del inglés *Contemporaneous AutoRegressive Moving Average*

[106]. En los trabajos [7], [100], [107], [108] son aplicados a series de velocidad de viento y potencia eólica. En todos estos casos se considera la transformada integral y análisis de estacionalidad a través del vector $\vec{\tau}_t$ de la ecuación (2.20). Este modelo multivariable consiste en diagonalizar la matriz de coeficientes A_i de la ecuación (2.19), lo que permite que sea posible ajustar un modelo ARMA(p,q) en forma independiente para cada sitio bajo análisis. Una ventaja radica en que la dependencia temporal puede ser caracterizada de mejor manera. La dependencia espacial es introducida a través del vector de ruido blanco, el cual se asume que se distribuye de acuerdo a $N(0, \Sigma)$. Al realizar esta suposición sólo se está modelando la correlación para el retardo cero, lo que sería una desventaja para este modelo. Se ha aplicado a series de velocidad de viento para frecuencias de muestreo de 1 hora. Sin embargo estas metodologías pueden ser aplicadas al recurso solar.

Otra ventaja que posee este modelo consiste en que no es necesario que las series de tiempo sean de igual longitud en la etapa de estimación de parámetros. Esta ventaja es analizada en [7], la que muestra que para situaciones en las cuales se considere la incorporación de nuevas plantas eólicas en donde no sea posible obtener mediciones de sincronizadas, es posible utilizar este modelo en lugar de un modelo VAR(p).

Una mejora al modelo CARMA es propuesta en [84] en donde se mantiene el supuesto de la diagonalización de la matriz de coeficientes autoregresivos pero además se consideran los coeficiente de correlación cruzada de algunos retardos en el proceso de muestreo. Este modelo fue utilizado en [9] para generar escenarios de integración de energía eólica con el propósito definir niveles de reserva a través de un modelo predespacho estocástico, una variación es realizada para incluir estacionalidades de las series de velocidad de viento.

Los modelos multivariables han sido aplicados a series de velocidad de viento, potencia eólica y potencia fotovoltaica. En la literatura se ha demostrado que a través de los preprocesamientos adecuados es posible caracterizar las series de tiempo utilizadas en cada artículo. Al igual que para los modelos univariados, se observa que existe una necesidad de quitar la estacionalidad, la cual depende específicamente de las características de la serie que se esté estudiando. Como principal desventaja se observa que la determinación de los parámetros es más compleja y puede ser prohibitivo utilizar estos modelos cuando la cantidad de sitios empieza a crecer considerablemente.

Teoría de Cópula

Como fue mencionado en la sección anterior la aplicación de teoría de Cópula para modelar recursos renovables en los sistemas eléctricos es relativamente nueva. De acuerdo a nuestro conocimiento, los primeros trabajos que consideran modelar dependencia temporal y espacial a través de teoría de Cópula son presentados en [75] y [76] en el año 2015. En [11] se utiliza la familia de Vine-Cópula para modelar velocidad de viento y potencia eólica, además se compara con el modelo multivariable. Se consideran series con frecuencias de muestreo de 1 hora y para caracterizar las variaciones estacionales se asume que cada estación se simula a través de 24 variables aleatorias. Se concluye que este modelo es capaz de caracterizar estructura de dependencia no lineal, las cuales se observan a través de la función de autocorrelación y diagrama de dispersión. Una desventaja se observa en el proceso de muestreo, debido a que el problema crece exponencialmente con el número de variables. Para evitar este problema los autores proponen truncar el algoritmo de muestreo para mitigar el efecto de la dimensión del modelo en la estimación de parámetros y en la generación de las series.

En [109] se utiliza la familia de diagonal-Cópula para modelar demanda, radiación y velocidad de viento. Es el único artículo que intenta considerar series de recursos y demanda en forma simultánea. Se obtienen buenos resultados en la representación estadísticas de las series originales. Además el artículo considera la transformación de radiación a potencia fotovoltaica.

Análisis de componentes principales más modelos ARMA

Debido a la complejidad de la estimación de parámetros en los modelos VAR(p) se propone en [110] utilizar Análisis de componentes principales (PCA⁹) que consiste en la aplicación de una transformación ortogonal

⁹ Del inglés *Principal Component Analysis*

para convertir un conjunto de series correlacionadas a otro conjunto de series no correlacionadas. La aplicación de componentes principales exige que el proceso sea Gaussiano y estacionario, de lo contrario luego de la transformación las series seguirán correlacionadas. Se propone quitar la estacionalidad de acuerdo a la ecuación (2.11) y convertir a proceso Gaussiano utilizando la transformada integral. Luego se aplica un modelo ARMA(p,q) a las componentes principales de cada sitio. El modelo puede ser aplicado tanto a series de potencia eólica así como a series de velocidad de viento y los autores exponen que puede extenderse a series de potencia fotovoltaica, demanda y radiación solar. Sin embargo, no existe un ejemplo en donde se muestre la factibilidad de ello.

Este modelo presenta grandes ventajas en situaciones donde se considera gran cantidad de variables, permitiendo disminuir la dimensión del problema. Los estadísticos y la función de densidad son modelados adecuadamente, por ello los autores concluyen que es posible utilizar el modelo para los distintos problemas que se resuelven en los sistemas eléctricos, tales como: flujos de potencia probabilístico, determinación de reserva, planificación de la transmisión, entre otros.

2.6 Análisis de modelos de generación de series sintéticas y problemas detectados

En la revisión bibliográfica realizada se observa que para generar series sintéticas se han utilizado principalmente los siguientes modelos matemáticos:

- Modelos temporales
 - Cadenas de Markov
 - ARIMA(p,d,q)
 - Teoría de Cópula
- Modelos espacio temporales
 - VAR(p)
 - Teoría de Cópula
 - Componentes principales más ARMA(p,q)

Exceptuando la teoría de Cópula, las metodologías expuestas tienen un factor común y consiste en que cada serie debe ser estacionaria y Gaussiana. Las propuestas para quitar la estacionalidad de las series son aceptadas en la comunidad científica pero no existe mayor análisis respecto al impacto en los resultados de la operación o planificación al considerar una u otra metodología. De la teoría de series de tiempo se puede inferir que la elección de algún determinado modelo es en función de las características de cada serie y no es posible realizar una generalización como es común encontrar en los artículos. Como ejemplo al comentario anterior, es común encontrar que las desventajas a los modelos ARMA(p,q) consisten en que no son capaces de representar la función de densidad de probabilidades de la serie real, debido a que el ruido blanco se distribuye con una función de densidad de probabilidad normal y las series de ERV no lo son. Esto no es correcto y nuevamente depende de la serie que se esté estudiando. Aplicando la transformación de Box-Cox o la transformación integral es posible obtener una serie que se distribuye en forma normal, para ajustar el modelo y luego realizar el proceso de transformación inverso. Este proceso también puede ser aplicado a una serie para luego aplicar una cadena de Markov, teóricamente se puede hacer pero no se ha reportado en la literatura.

Los modelos espacio temporales han sido principalmente del tipo VAR. Con estos modelos es posible caracterizar adecuadamente la función de densidad y los estadísticos, además la obtención de los parámetros se hace a través de minimización de mínimos cuadrados, facilitando su aplicación. No obstante, tienen desventajas en comparación con la teoría de Cópula cuando las series bajo análisis representan estructura de correlación más compleja. La teoría de Cópula matemáticamente es más compleja, tanto en su formulación, estimación de parámetros y proceso de simulación [94]. Además se debe escoger una familia de funciones de Cópula y esta tarea puede ser difícil ya que no existe una metodología exacta.

No se han aplicado modelos VARMA ni VARMA-GARCH a series de recursos renovables. Estos últimos podrían servir para modelar series con mayor volatilidad. Tampoco se han analizados modelos espectrales que consideren la dependencia espacio temporal. Sin embargo, el conjunto de modelos ya desarrollados

entrega suficiente información para caracterizar adecuadamente los recursos en los problemas de sistemas eléctricos en los cuales se cuenta con mediciones.

La inteligencia computacional ha sido utilizada para generar series sintéticas de radiación y potencia fotovoltaica, pero a la fecha no han sido aplicados en otro tipo de series de tiempo de ERV. Sin embargo, no implica que no puedan ser aplicados.

A diferencia de los modelos de pronósticos, en los modelos para generar series sintéticas no se han utilizado los mismos indicadores para analizar la bondad de cada modelo. Es usual un análisis cuantitativo de los estadísticos y un análisis cualitativo de la semejanza de las funciones de autocorrelación simple y parcial. Cuando se analizan series que contienen estacionalidades marcadas se han graficado las medias para comprobar las bondades de los modelos, pero no existen otros estadísticos para verificar contribuciones energéticas estacionales. Esta característica es de vital importancia para modelos de operación de largo plazo.

2.7 Resumen y conclusiones del estado del arte

En este capítulo se ha presentado la evolución de las ERV (energía eólica y solar) a nivel mundial y nacional quedando de manifiesto que el escenario de variabilidad e incertidumbre de los recursos será aún mayor en el futuro. De los distintos problemas que son considerados en los sistemas eléctricos, las series sintéticas se han utilizado en: procesos de simulación de Monte Carlo para resolver problemas tales como flujos de potencia probabilístico, predespacho estocástico, estudios de confiabilidad y problemas de planificación. Para problemas de operación en tiempo real no se consideran las series sintéticas.

En las Figuras 2.12 -2.15 se muestran los esquemas de los distintos modelos utilizados para generar series sintéticas según los recursos y tipo de correlación que consideran en su formulación. Estos esquemas permiten conocer las estructuras que pueden ser aplicadas en función de los recursos. Para un sistema en presencia de plantas solares fotovoltaicas un modelo VAR(p) y teoría de Cópula pueden ser empleados, siendo dos artículos los que han mostrado resultados experimentales de la utilización de ellos [86][109]. Todos los modelos desarrollados y aplicados en la literatura son estadísticos, necesitan de registros históricos para ser ajustados. Se observa que tienen gran aceptación, en especial los modelos ARMA y VAR(p). Esto ocurre porque son conocidos, su teoría ha sido validada, la estimación de los parámetros es relativamente sencilla, pueden realizarse simulaciones desde un computador y finalmente pueden ser aplicados por personas que no sean especialistas en meteorología [46]. Una desventaja de estos modelos consiste en que el requisito necesario para aplicarlos es que la serie sea estacionaria y esto no ocurre con series de ERV. Sin embargo, a través de distintos procesamientos previos es posible cumplir con estos requisitos y aplicar satisfactoriamente estos modelos [105]. El procesamiento previo más utilizado para lidiar con la no normalidad de las series de ERV consiste en la transformada integral.

Las principales características que los modelos de series sintéticas buscan replicar son: función de autocorrelación simple, función de autocorrelación parcial, función de densidad de probabilidad y función de correlación cruzada. Los modelos desarrollados en la literatura muestran que son capaces de rescatar estas características de los recursos siempre y cuando se haga un análisis y se apliquen las transformaciones pertinentes a las series antes de realizar los ajustes del modelo. No realizar la tarea anterior lleva a conclusiones no generales y erróneas. En su mayoría, los artículos no presentan novedad en la estructura matemática de los modelos, más bien el aporte consiste en utilizar teoría de series de tiempo, análisis probabilístico para modelar la variabilidad e incertidumbre de los recursos renovables en distintos sitios y con distintos objetivos. Es necesario decir que las series sintéticas solo reproducen estadísticas similares a las existentes y no pueden generar información extra.

Dada la gran cantidad de artículos que han sido escritos, es una tarea compleja escoger un modelo dado un problema de sistema eléctrico, ello justifica la propuesta metodológica de este trabajo. El último trabajo que realiza una revisión bibliográfica y compara distintos modelos univariados fue realizado en [77], en donde se comparan las funciones de autocorrelación, respuesta en frecuencia y funciones de densidad. Para el caso de los modelos multivariados no existe un artículo que realice estas comparaciones.

Una descripción general del campo de acción de las series sintéticas en los SEP puede observarse en la Figura 2.16. Para los problemas de mediano y largo plazo las series pueden tener una duración mayor a 1 año, lo que implica que deben ser capaces de modelar adecuadamente las variaciones estacionales con sus respectivas funciones de densidad, con ello se tendrá una adecuada estimación de la energía que suministrarán las ERV [33]. Para los problemas de corto plazo y muy corto plazo es necesario caracterizar adecuadamente la función de autocorrelación y la función de densidad de probabilidades. La primera de estas funciones representa la persistencia de los recursos y la segunda la incertidumbre. Estos elementos son importantes, porque se aísla el problema de SEP en función de los horizontes de evaluación y será ese criterio el que permite sistematizar y buscar un modelo adecuado. Una observación respecto a la utilización de series sintéticas en el corto plazo, cuando se utilizan en estos problemas es para generar escenarios a través de una simulación de Montecarlo y luego reducir para finalmente solucionar el problema que se esté analizando.

En las tablas Tabla 2.4 y 2.5 se muestra un resumen con los modelos propuestos en la literatura, separadas por tipo de dependencia que consideran. Este resumen presentado en tablas forma parte de la metodología propuesta que se presenta en el siguiente capítulo. Brevemente se comenta el tipo de modelo, a qué tipo de recurso ha sido aplicada, ventajas, desventajas y fuente. La columna de aplicación en SEP comenta si en alguno de los artículos se presenta alguna aplicación específica a un problema de SEP, y la ausencia de comentarios en la celda no implica que el modelo no pueda ser usado en algún problema específico de SEP. Además en el Anexo 4 se presenta el listado de modelos según artículo y recurso, las Tabla 10.1-12.4 también serán utilizadas como parte de la metodología de selección.

Para este trabajo de tesis, dada la revisión bibliográfica presentada en este capítulo, se desarrollan modelos ARMA (p,q), Cadenas de Markov y VAR(p). La elección de estos modelos se debe a que son los más aplicados en la generación de series sintéticas. Se realiza especial énfasis se ha realizado en los modelos VAR(p) ya que, como se observa en la Figura 2.17, cuando se trata de modelar la correlación espacio temporal, esta estructura matemática es la que más se ha utilizado, debido a que a través de preprocesamientos previos (ver Figura 2.18) ha sido posible convertir series de ERV a series que cumplan con los requisitos para aplicar adecuadamente estos modelos.

Para los modelos ARMA (p,q) y VAR(p) se considera la transformada integral para convertir los procesos a Gaussianos, dado que es la técnica más usada. Esta técnica requiere la estimación de la función de distribución de probabilidades, la cual se realiza utilizando Kernel Gaussianos. La propuesta metodológica de selección será validada con ellos, sin embargo se puede aplicar para todas los modelos propuestos en la literatura.

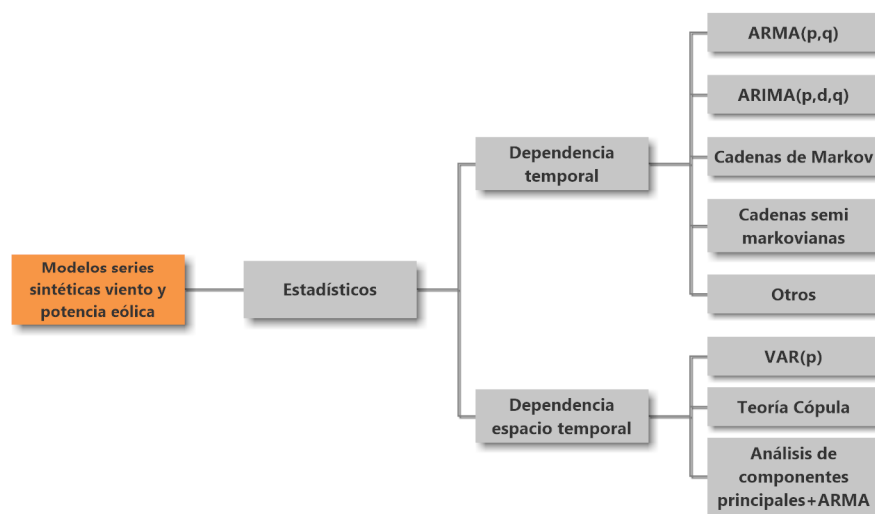


Figura 2.12. Esquema de modelos para generar series sintéticas de velocidad de viento y potencia eólica

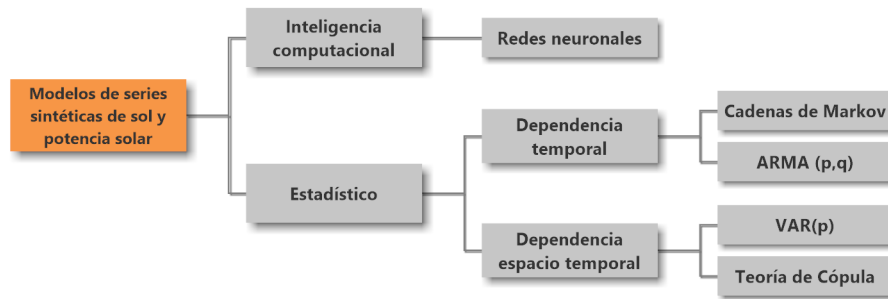


Figura 2.13. Esquema de modelos para generar series sintéticas de radiación y potencia fotovoltaica



Figura 2.14. Esquema de modelos para generar series sintéticas que consideran recurso solar y eólico



Figura 2.15. Esquema de modelos para generar series sintéticas que consideran como variable radiación, velocidad de viento y demanda

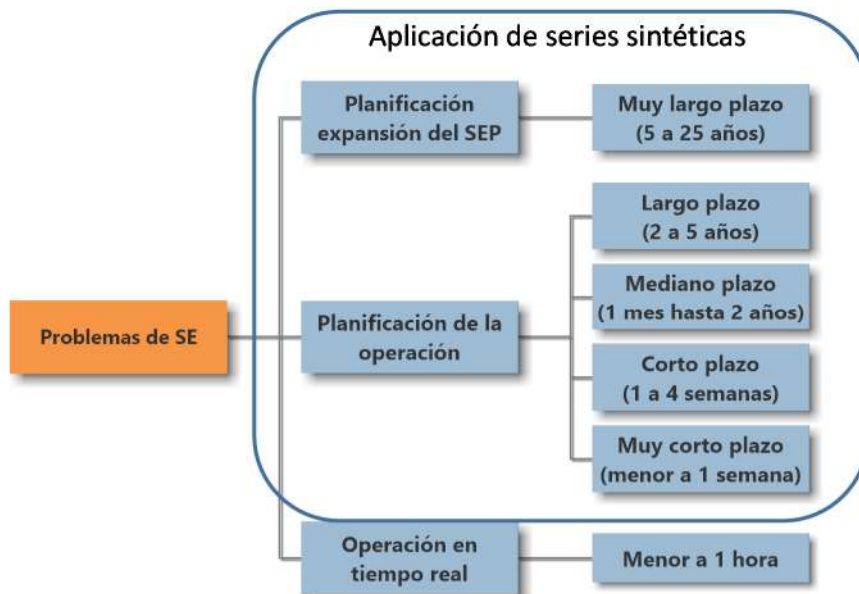


Figura 2.16. Esquema de problemas de SEP y sus respectivos horizontes de tiempo

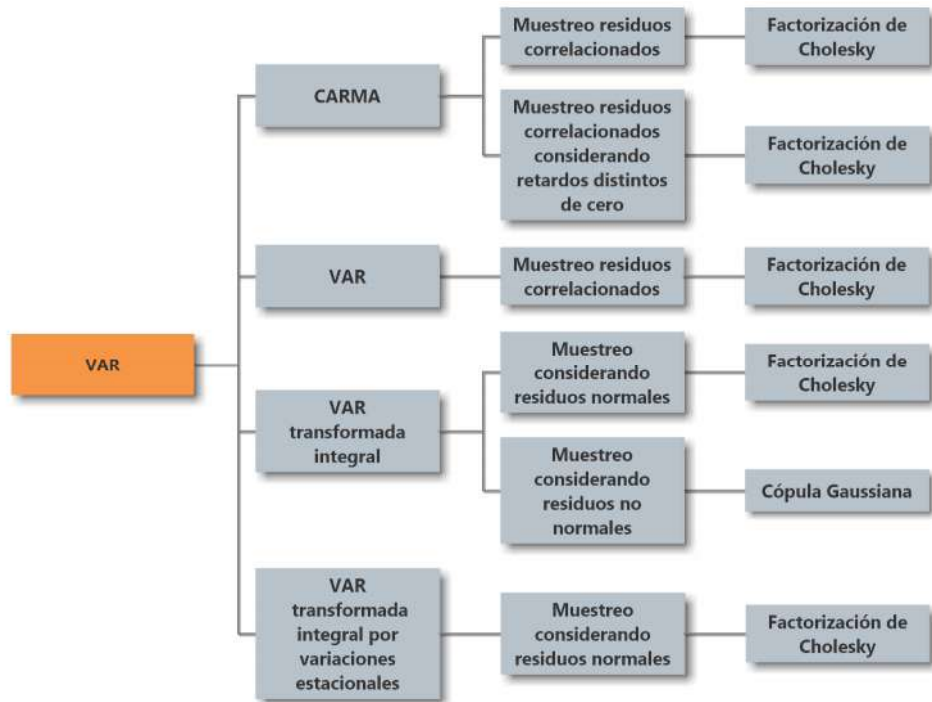


Figura 2.17. Esquema con distintas variantes de modelos VAR reportados en la literatura.

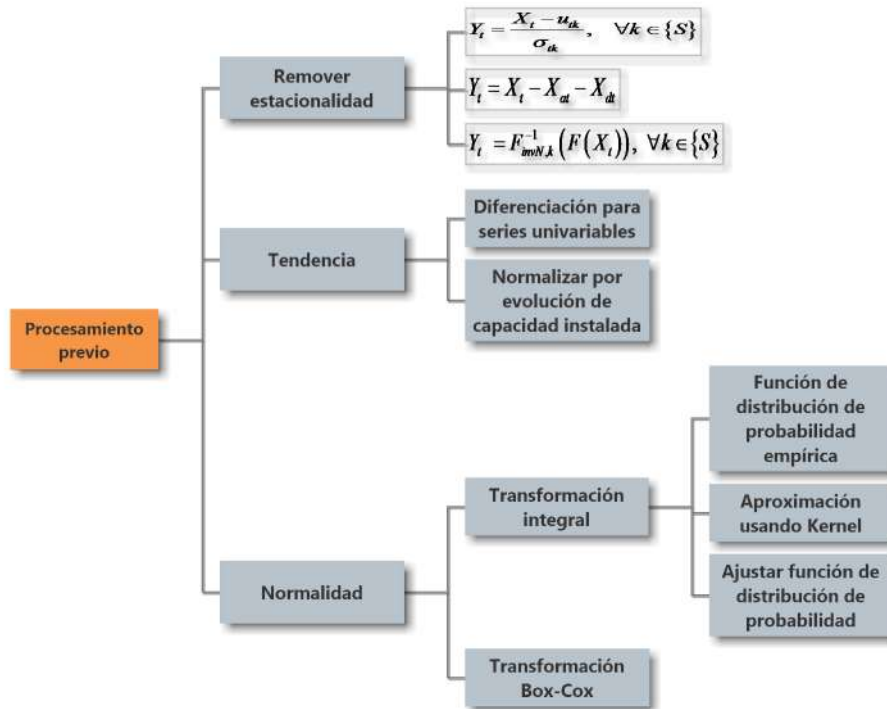


Figura 2.18. Procesamientos previos realizados para remover tendencia, estacionalidad y transformar las series de ERV a normales

Tabla 2.4. Resumen de modelos para generación de series sintéticas con dependencia temporal de acuerdo a la literatura (Fuente: elaboración propia)

N°	Modelo	Variable a modelar	Frecuencia muestreo	Requisitos	Aplicación a SEP	Ventajas	Desventaja	Fuente
1	Cadenas de Markov homogénea 1°, 2° y 3° orden	Velocidad Potencia eólica Índice de Claridad Radiación	30-60 min 15 min 1 min	Serie estacionaria Longitud de medición igual o superior a 1 año	Modelo utilizado para estimar sistema de almacenamiento Planificación de la generación	Buena reproducción de función de distribución de probabilidades y principales estadísticos Fácil aplicación	Puede no representar adecuadamente función de autocorrelación y necesita de una gran cantidad de parámetros	[77][48][50] [53][111] [58][60] [54][61]
2	Cadenas semi Markoviana 1°, 2° y 3° orden	Velocidad	10 min	Serie estacionaria Longitud de medición igual o superior a 1 año	No se ha reportado	Buena reproducción de función de distribución de probabilidades y principales estadísticos	Puede no representar adecuadamente función de autocorrelación y necesita de una gran cantidad de parámetros	[112]
3	ARMA(p,q) ARIMA(p,d,q)	Velocidad Potencia eólica Índice de Claridad radiación	60 min	Serie estacionaria Longitud de medición igual o superior a 1 año	Estudios de confiabilidad en SEP Coordinación Hidrotérmica Estudios de impacto en SEP	Buena reproducción de Función de autocorrelación, Función de densidad de probabilidad y Estadísticos	Muy dependiente de los parámetros del modelo Se necesita de transformaciones para aplicar correctamente	[74][113][68] [34][82]
4	ARMA (p,q) GARCH (a,b)	Velocidad Radiación	60 min	Serie estacionaria Longitud de medición igual o superior a 1 año	No se ha reportado	Buena reproducción de Función de autocorrelación, Función de densidad de probabilidad y Estadísticos La volatilidad del recurso queda mejor representada	Muy dependiente de los parámetros del modelo Se necesita de transformaciones para aplicar correctamente	[77]

Tabla 2.5. Resumen de modelos para generación de series sintéticas con dependencia espacio-temporal de acuerdo a la literatura (Fuente: elaboración propia)

N°	Modelo	Variable a modelar	Tiempo de muestreo	Requisitos	Aplicación a SEP	Ventajas	Desventaja	Fuente
1	VAR(p)	Velocidad Radiación	60 min	Serie estacionaria Longitud de medición igual o superior a 1 año	Ejemplo sencillo para definir niveles de transmisión Generación de escenarios para estudios de integración de ERNC en SEP	Buena reproducción de los estadísticos, función de autocorrelación, función de autocorrelación cruzada y función de densidad de probabilidades.	Todas las series deben tener igual longitud La determinación de parámetros cuando son muchas series es un proceso complejo Complejidad de agregar nuevas plantas donde no existan mediciones Cuando la estructura de la dependencia es muy no lineal puede haber problemas en la función de autocorrelación cruzada Necesita transformación para convertir a proceso Gaussiano	[114][15] [104] [100]
2	CARMA(p,q)	Velocidad	60 min	Serie estacionaria Longitud de medición igual o superior a 1 año. Las series pueden ser de longitud distinta	Generación de escenarios para despacho económico	Buena reproducción de los estadísticos, función de autocorrelación y coeficiente de correlación cruzada para retardo igual a cero Permite ajustar modelos ARMA de diferente orden por sitio	No se considera coeficientes de correlación cruzados para retardos mayores a cero Muy dependiente de los parámetros del modelo Necesita transformación para convertir a proceso Gaussiano	[7], [100], [107] [116][108]
3	VAR(p) matriz diagonalizada	Velocidad	60 min	Serie estacionaria Longitud de medición igual o superior a 1 año. Las series pueden ser de longitud distinta	Generación de escenarios para resolver predespacho estocástico y determinar reservas Estudio de confiabilidad Generación de escenarios para estudiar impacto en SEP por integración ERNC	Buena reproducción de los estadísticos, función de autocorrelación cruzada y función de densidad de probabilidades. Permite ajustar modelos ARMA de diferente orden por sitio	Muy dependiente de los parámetros del modelo Cuando la estructura de la dependencia es muy no lineal puede haber problemas en la función de autocorrelación cruzada Necesita transformación para convertir a proceso Gaussiano	[117][9] [103][7]
4	PC+ARMA(p,q)	Velocidad o potencia eólica	60 min	Serie estacionaria Longitud de medición igual o superior a 1 año	Estudio de confiabilidad	Buena reproducción de estadísticos Dependencia temporal y espacial Componentes principales descomponen dependencia espacial y se aplica diferentes modelos	Necesita transformación para convertir a proceso Gaussiano	[110]
5	Teoría de Copula (Vine-Copula)	Velocidad o potencia eólica	60 min	Longitud medición igual o superior a 1 año	Análisis de sistemas de almacenamiento en SEP	Buena representación de los estadísticos, las funciones de autocorrelación y función de autocorrelación. Puede representar estructura de dependencia no lineales	Mayor complejidad en el proceso de simulación Escoger la familia de funciones de Cópula es un proceso de prueba y error	[11]
6	Teoría de Copula (Diagonal band Copula)	Velocidad Radiación Demanda	60 min	Longitud medición igual o superior a 1 año	No se ha reportado	Buena representación de los estadísticos, las funciones de autocorrelación y función de autocorrelación. Puede representar estructura de dependencia no lineales	Mayor complejidad en el proceso de simulación Escoger la familia de funciones de Cópula es un proceso de prueba y error	[109]

3 Propuesta metodológica para selección de modelos de generación de series sintéticas

3.1 Estructura general de la metodología propuesta

Uno de los principales objetivos de este trabajo de investigación consiste en sistematizar la selección de modelos para generar series sintéticas en función del problema de SEP que se desea resolver. En esta sección se describe en términos generales la propuesta metodológica. En la Figura 3.1 se muestra un diagrama de flujo general de la metodología, en cada etapa existen una serie de procesos que deben ser realizados para cumplir con el objetivo, cada uno de ellos será detallado en las secciones siguientes de este capítulo.

La metodología busca sistematizar el proceso de toma de decisiones de un experto en modelos estocásticos y un especialista en SEP para escoger un modelo adecuado teniendo en cuenta los objetivos de los estudios en SEP así como la disponibilidad y características de las series de tiempo asociadas a los recursos. Para ello, existen etapas cuyo propósito es obtener información de los problemas de SEP y las series de tiempo de recursos, y otras etapas en donde se procesa la información obtenida para que finalmente se recomiende un conjunto de modelos reportados en la literatura. La información es obtenida a través de una serie de preguntas que deben ser contestadas. Las respuestas se encuentran acotadas dentro de un conjunto factible, las cuales se han definido en función de las posibles situaciones que pudieran enfrentar cuando se analiza la integración de ERV en los SEP.

A continuación se describe en términos generales el diagrama de flujo:

1. Definir problema de SEP: el objetivo de esta etapa consiste en especificar si el problema será de planificación de la operación o planificación de la inversión y obtener información de las consideraciones que se están haciendo respecto a la modelación del sistema eléctrico ¿Se considera o no el sistema de transmisión? ¿Cuál es el horizonte de simulación? ¿Simulación cronológica o a través de curvas de duración? entre otras. En esta etapa se realiza una serie de preguntas para obtener la información que se traspa a las etapas 3, 5 y 6.
2. Definir tipo de serie del recurso solar y/o eólico disponible: El objetivo de esta etapa consiste en obtener información del tipo de dato al cual se tiene acceso. Como fuente de información se han considerado las mediciones, reportes de operadores de red y series de tiempo generadas a través de modelos fenomenológicos. También se busca obtener información de los tiempos de muestreo de cada una de las series.
3. ¿Existe información necesaria para desarrollar el estudio? : El objetivo de esta etapa de decisión consiste en coordinar las dos etapas anteriores respecto a los tiempos mínimos de simulación. Si, por ejemplo, en la etapa 1 se quiere hacer un predespacho cuyo tiempo mínimo de resolución es 30 minutos y las series de tiempo tienen un tiempo mínimo de 60 minutos, se debe ajustar el problema de SEP para que tenga el mismo tiempo de resolución que las series de tiempo de los recursos. En este trabajo no se han revisado modelos que buscan obtener información para tiempos inferiores a los que se dispone. Por ello, se considera que estos dos elementos estén sincronizados.
4. Análisis estadísticos de las series del recurso solar y/o eólico: El objetivo es poder obtener información de las características estadísticas de la serie, tales como: función de distribución, análisis de correlación y variaciones estacionales anuales y diarias. A través de un análisis exploratorio de los datos y pruebas estadísticas se responden las preguntas a esta etapa. También se definen transformaciones y preprocesamientos adecuados que pueden ser utilizados para poder aplicar los modelos.
5. Definir conjunto de modelos candidatos: El objetivo es utilizar la información de las etapas 3 y 4 para definir un conjunto de modelos de acuerdo al estado del arte. A través de una matriz desarrollada en base al estado del arte, se buscan todos aquellos modelos que cumplan con las condiciones del problema de SEP.
6. Selección del modelo: El objetivo consiste primero en ajustar los parámetros y verificar la bondad de los modelos candidatos de la etapa 5. A continuación, para cuando exista más de un modelo

candidato, se calculan distintos indicadores para poder comparar y así, en función de los indicadores y la información de la etapa 1, tomar la decisión del modelo adecuado para generar series sintéticas.

De acuerdo a la Figura 2.13 se han utilizado modelos VAR(p) y teoría de Cópula cuando se consideran series de radiación y de velocidad de viento. Ante esta situación, se podría concluir que no existen muchos modelos para escoger, sin embargo, para el caso de los modelos VAR(p) se han presentados distintas metodologías de procesamiento previo para poder ajustar y generar series sintéticas (ver capítulo 2, sección 6). Como se comentó en el capítulo anterior el modelo más referenciado es el VAR(p) y en la etapa de selección se propone una metodología que permita escoger el orden, tipo de procesamiento previo y técnica de muestreo, siendo un aporte a la sistematización de selección de modelos VAR(p), tema que no ha sido expuesto en la literatura existente.

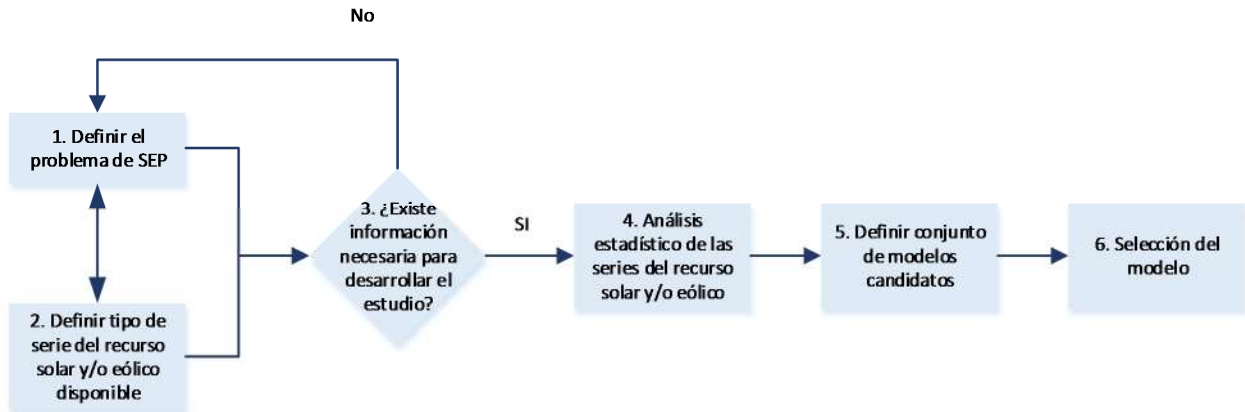


Figura 3.1. Diagrama de flujo general de metodología propuesta

3.2 Descripción de las componentes de la metodología propuesta

3.2.1 Definir problema de SEP

En la Figura 3.2 se puede observar un esquema general de la etapa 1. Dentro de esta etapa se busca definir el tipo de problema que se desea resolver, tipos de modelos de SEP considerados, tiempos de resolución, etc. A continuación se detalla cada sub etapa definiendo objetivos y razones por las cuales se incluyen en la metodología.

Tipo de problema:

En esta subetapa se desea definir si el problema planteado consiste en planificar la operación del sistema eléctrico o planificar la inversión en generación o transmisión del SEP. De acuerdo al estado del arte, cuando el análisis considera la simulación de largo plazo es importante que las series sintéticas puedan representar adecuadamente las variaciones estacionales, dado que influirá en la cantidad de energía que generarán las centrales ERV. Por otro lado, para el caso de planificación de corto plazo, las series sintéticas son utilizadas para generar escenarios y se buscará que la dependencia temporal sea muy bien caracterizada. Las preguntas que deben ser contestadas consideran la división en función del horizonte de tiempo de los problemas de planificación de la operación e inversión que se muestran en la Tabla 2.3.

Tipo de simulación:

En esta subetapa se desea definir el tipo de simulación que será utilizado para el problema de SEP que se desea solucionar. En los problemas de planificación de la operación e inversión se utilizan generalmente dos tipos simulación. En la primera se considera cómo cambian las distintas variables a través del tiempo. Por ejemplo, en un predespacho lo que se desea es poder definir el conjunto de unidades que estarán prendidas y apagadas para un horizonte de tiempo. Por otro lado en la segunda simulación es usual modelar la demanda a través de una curva de duración, como es el caso de la planificación de mediano plazo en

sistemas hidrotérmicos [118]. La elección del tipo de simulación permite definir el tipo de característica que la serie sintética debe emular, tal y como fue detallado en el párrafo anterior.

Tipo de modelo del sistema eléctrico:

El objetivo de esta subsección es definir el nivel de detalle con que se modelará el sistema de transmisión. Se consideran dos opciones: un sistema uninodal o multinodal. La importancia de esta subetapa radica en que se podría simplificar la elección de un modelo: si el estudio considera una simulación uninodal, es posible ajustar un modelo para la serie agregada de ERV, en vez de realizar un ajuste para cada central del SE. Ajustar un modelo para un sistema agregado implica que se podrían considerar modelos que solo consideran dependencia temporal, facilitando el proceso de ajuste y generación de serie sintética [46]

Horizonte de evaluación:

El objetivo de esta subetapa consiste en definir el horizonte de evaluación que se desea simular. En la Figura 3.2 se muestran las posibles opciones que se pueden tener en forma general. Sin embargo, para la selección son dos los rangos que interesan: menor a un año y mayor a un año. Esto se relaciona en el contexto de aplicación de series sintéticas según se observó en la Figura 2.16. Lo anterior implica el tipo de transformación, preprocesamiento e indicadores que deben ser considerados en la toma de decisión.

Tiempo mínimo de simulación:

Se busca definir el tiempo mínimo de simulación, pudiendo ser del orden de una hora para un problema de predespacho o semanas para un problema de coordinación hidrotérmica. La información obtenida en esta subsección sirve como dato de entrada para la etapa de selección del modelo.

Energías renovables:

El objetivo de esta subetapa es definir cuál será la situación de las centrales ERV que se desean considerar en el problema de SEP. En la práctica se tiene el análisis para sistemas con plantas operando y futuras plantas. Conocer el tipo de centrales que serán analizados se relaciona con las series de recursos que se disponen. Algunos modelos en la literatura exigen que los datos se encuentren sincronizados y otros realizan algunas simplificaciones para poder analizar situaciones en las cuales se consideran series de distinta longitud o en lugares en los cuales no se dispone información de la serie [7].

En el caso de la energía eólica el proceso de conversión entre velocidad de viento y potencia es no lineal y se ve influenciada por diversos factores cuando se desea estimar la potencia eólica de un parque en función de los datos de una turbina [33]. Para problemas de largo plazo en donde se evalúa la penetración de distintas centrales es usual considerar series de velocidades de viento en vez de potencia. Cuando el horizonte es menor y se dispone de series de potencia es preferible utilizar las series de potencia porque se evita el error de conversión [53].

Cuestionario

Para poder obtener la información deseada de cada una de las subetapas mencionadas se deben contestar las siguientes preguntas con sus respectivas alternativas y almacenar las respuestas. Para esta etapa se tiene un conjunto de 6 preguntas las que se detallan a continuación, donde para cada pregunta se considera una alternativa.

1. ¿Qué tipo de problema de sistema eléctrico se desea resolver?

Las respuestas válidas para esta pregunta son:

- a) Planificación de la operación largo plazo
- b) Planificación de la operación mediano plazo
- c) Planificación de la operación corto plazo
- d) Planificación de la operación de muy corto plazo
- e) Planificación de la inversión de generación
- f) Planificación de la inversión en transmisión

2. ¿Qué tipo de simulación se utilizará?

Las respuestas válidas para esta pregunta son:

- a) Simulación cronológica
- b) Simulación mediante curva de duración

3. ¿Qué tipo de modelo del sistema eléctrico es considerado?

Las respuestas válidas para esta pregunta son:

- a) Sistema uninodal
- b) Sistema multinodal

4. ¿Cuál es el orden del horizonte de evaluación del problema de SEP?

Las respuestas válidas para esta pregunta son:

- a) Minutos
- b) Horas
- c) Semanas
- d) Meses
- e) Años
- f) días

5. ¿Cuál es el tiempo mínimo de simulación?

Las respuestas válidas a esta pregunta son:

- a) Segundo
- b) Minuto
- c) Hora (considera tiempos t , tal que $1 \text{ min} < t \leq 60 \text{ minutos}$)
- d) Semana
- e) Mes

6. ¿Cuáles son los requerimientos de los modelos para centrales renovables?

Las respuestas válidas para esta pregunta son:

- a) Modelar sólo plantas existentes
- b) Modelar distintos niveles de penetración de energías renovables
- c) Modelar plantas existentes y plantas nuevas
- d) Modelar plantas existentes más distintos niveles de penetración de energías renovables

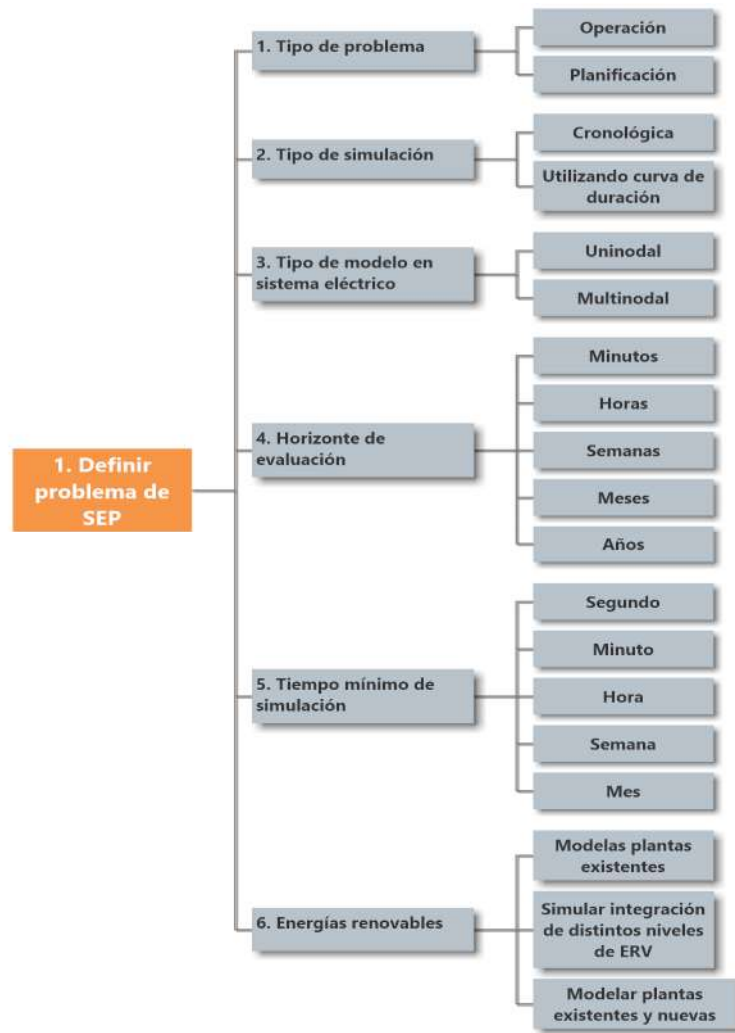


Figura 3.2. Mapa conceptual etapa 1: Definir problema de SEP

En el diagrama mostrado en la Figura 3.3 se muestran cómo se relaciona la información obtenida con las siguientes etapas. Se ha considerado utilizar una nomenclatura para referenciar a cada pregunta y respuesta de la siguiente forma: EXPY, donde las letras X e Y corresponden al número de etapa y número de pregunta respectivamente. Así, el código E1P2 indica que etapa 1 y pregunta 2.

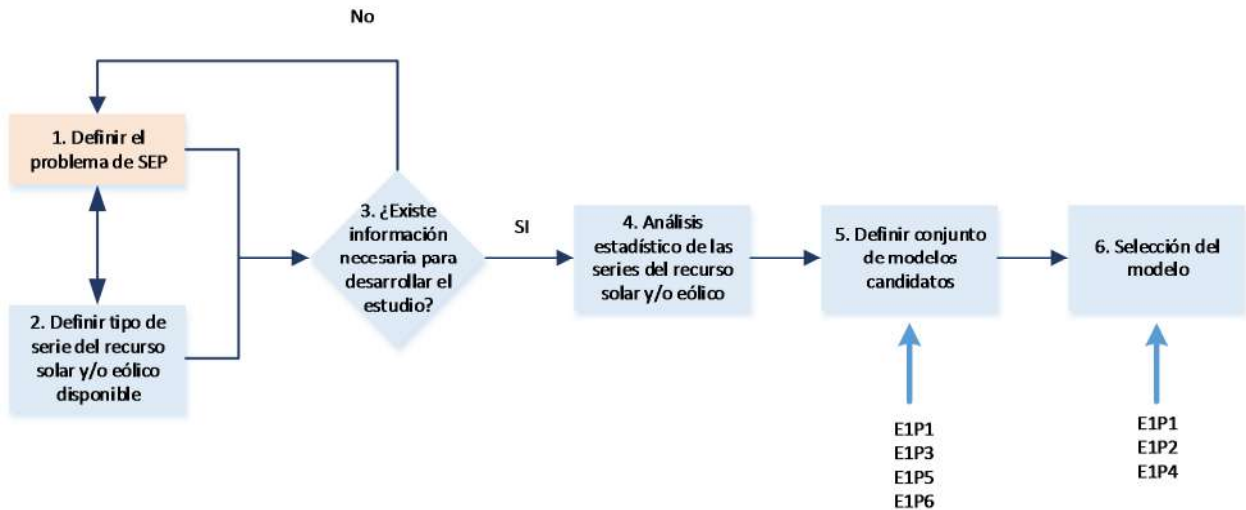


Figura 3.3. Relación entre preguntas de la etapa 1 con el resto de las etapas.

3.2.2 Definir tipo de serie del recurso solar y/o eólico disponible

El esquema general de esta etapa es presentado en la Figura 3.4. El objetivo principal de esta etapa es definir los tipos de series de recursos disponibles. En el estado del arte se observó que todos los modelos para generar series sintéticas son del tipo estadísticos y se ajustan dado un conjunto de mediciones. En grandes estudios de integración realizados en distintas partes del mundo, si no se poseen mediciones en los lugares de interés, se utilizan datos generados a través de modelos fenomenológicos [119], por lo tanto estos datos también pueden ser considerados datos de entrada para ajustar modelos de series sintéticas [9]. Para cada una de las series de recursos se pueden presentar distintas situaciones, las cuales quedan representadas en las Figura 3.5 y 3.6. Reconocer adecuadamente la naturaleza de las distintas series de tiempo permite realizar un filtrado de modelos, facilitando la tarea de selección.

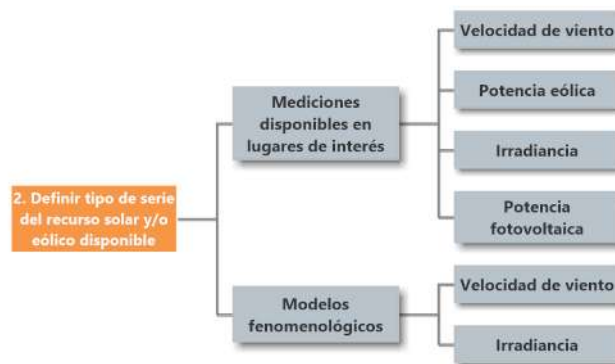


Figura 3.4. Esquema general etapa 2 de la metodología propuesta

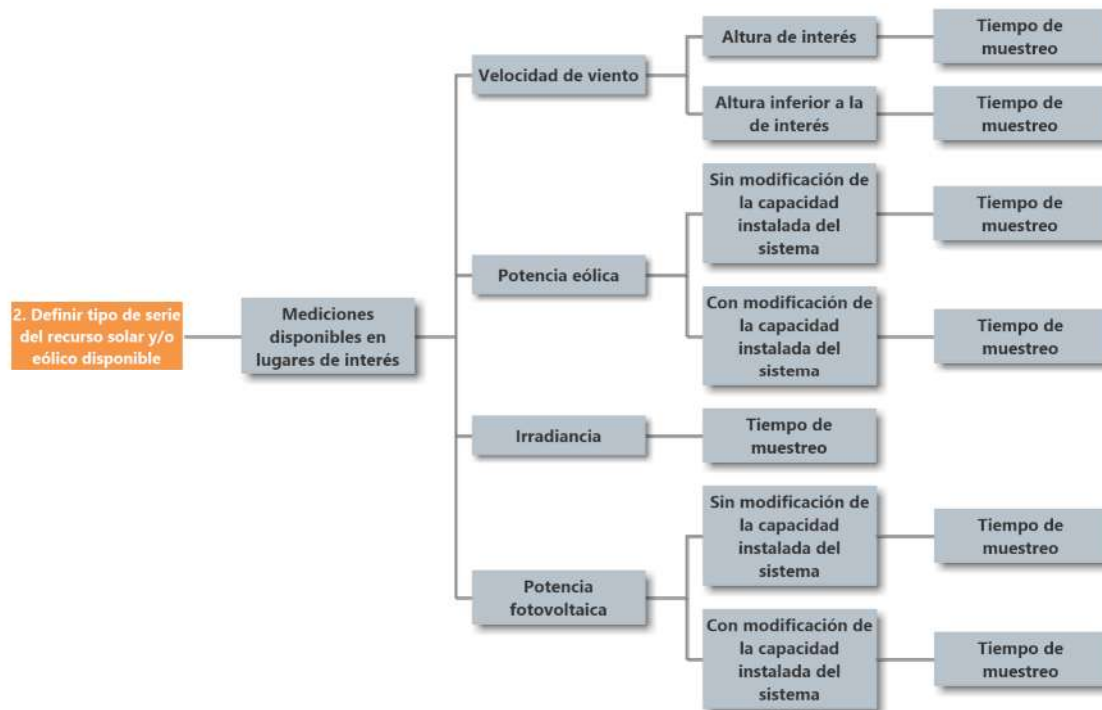


Figura 3.5. Detalle de subsección Mediciones disponibles en lugares de interés

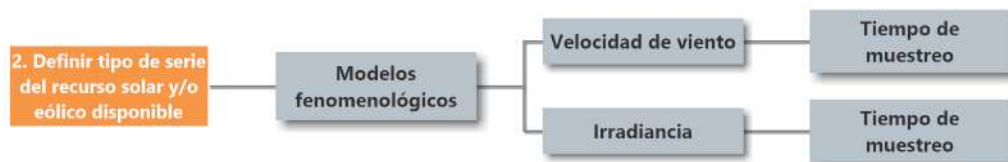


Figura 3.6. Detalle de subsección modelos fenomenológicos.

Cuestionario

La información en esta etapa es obtenida a través de las respuestas a las siguientes preguntas:

1. ¿Cómo se generaron las series con registros de los recursos renovables?

Las respuestas válidas a esta pregunta son:

- a) Mediciones del recurso
- b) Obtención de datos a través de modelos fenomenológicos

Esta pregunta puede aceptar como respuesta correcta ambas alternativas.

2. ¿Cuál es el recurso del que se tiene información?

Las respuestas válidas a esta pregunta son:

- a) Velocidad de viento a la altura de las turbinas eólicas
- b) Velocidad de viento a una altura inferior de las turbinas eólicas
- c) Potencia eólica de cada parque en el sistema eléctrico
- d) Potencia eólica agregada por región en el sistema eléctrico
- e) Radiación solar
- f) Potencia fotovoltaica de cada parque en el sistema eléctrico
- g) Potencia fotovoltaica agregada por región en el sistema eléctrico

La respuesta a esta pregunta consiste en marcar todas las posibles opciones que se encuentran disponibles. Independiente de la fuente de información a la cual se acceda, para ambos casos es necesario saber el tiempo mínimo de registro que arroja el modelo. En el caso de las mediciones es usual que los equipos tengan tiempo de muestreo del orden de los minutos o segundos pero se registren sólo los promedios para intervalos mayores de tiempo.

3. ¿Cuál es la resolución temporal de los registros?

Las posibles respuestas para esta pregunta son:

- a) Segundo
- b) Minuto
- c) Hora (considera tiempos t , tal que $1 \text{ min} < t \leq 60 \text{ minutos}$)
- d) Semana
- e) Mes

La respuesta a esta pregunta puede ser más de una alternativa. Las respuestas a todas las preguntas alimentan con información las etapas 5 y 6 como se observa en la Figura 3.7.

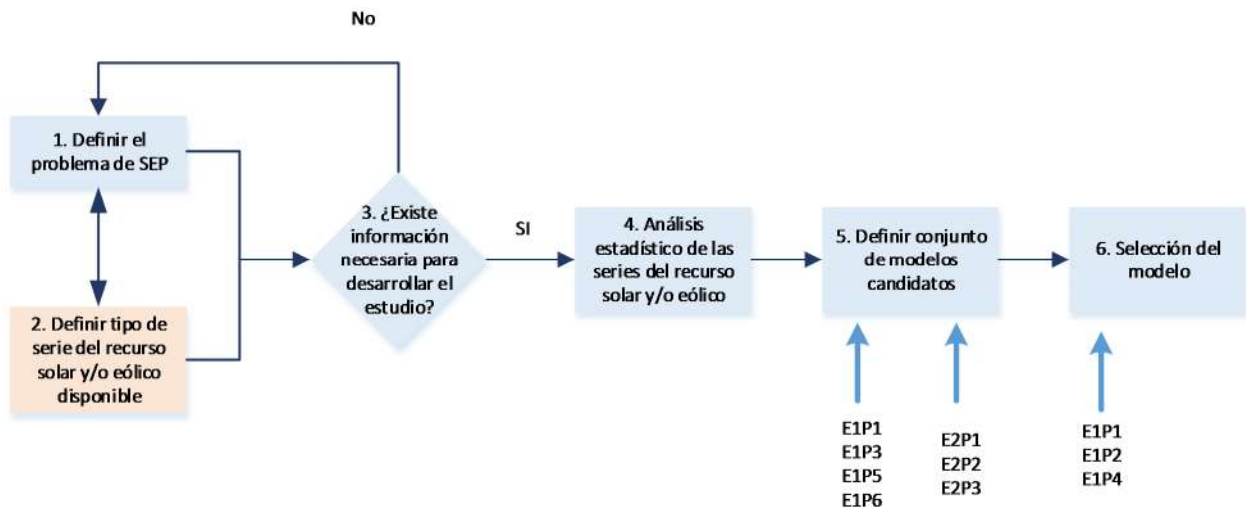


Figura 3.7. Relación entre preguntas de la etapa 2 con el resto de las etapas de la metodología propuesta.

3.2.3 Etapa de decisión: ¿Existe información necesaria para desarrollar el estudio?

La etapa de decisión consiste en compatibilizar el tiempo mínimo de resolución del SEP junto al tiempo mínimo de muestreo. Se pasa a la siguiente etapa siempre y cuando el tiempo de resolución de los registros de series de tiempo sea igual o menor que el tiempo mínimo de simulación. Existen algunos modelos que buscan interpolar cuando no se disponga de la resolución deseada. En este trabajo de investigación, no se consideran aproximaciones de los recursos para tiempos menores a los disponibles. Por ello, se asume que el problema de SEP debe concordar con la disponibilidad de las series de recursos.

3.2.4 Análisis estadístico de las series del recurso solar y/o eólico

El objetivo de esta etapa consiste en realizar una descripción estadística de la serie de tiempo asociada a los recursos obtenidos en la etapa 2. Se busca determinar la existencia de tendencias, estacionalidades, correlaciones temporales y espaciales en las series de tiempo de recursos. Esto permite acotar el conjunto de modelos que podrían ser utilizados en la etapa de selección, así como definir transformaciones y procesamiento previos adecuados para ajustar los modelos en caso de ser necesarios.

Para cumplir con el objetivo de esta etapa, se ha considerado utilizar análisis exploratorio de datos [120], el cual consiste en observar una serie de gráficos e inferir respecto a las características de las series de tiempo

de las ERV. Además, si a través del análisis exploratorio de datos no es posible concluir respecto a las características de las series de tiempo de las ERV, entonces se pueden utilizar pruebas de hipótesis estadísticas que permitan concluir cuando existan dudas. El esquema general es mostrado en la Figura 3.8. Esta etapa contiene cuatro subetapa, cada una de ellos es explicada a continuación.

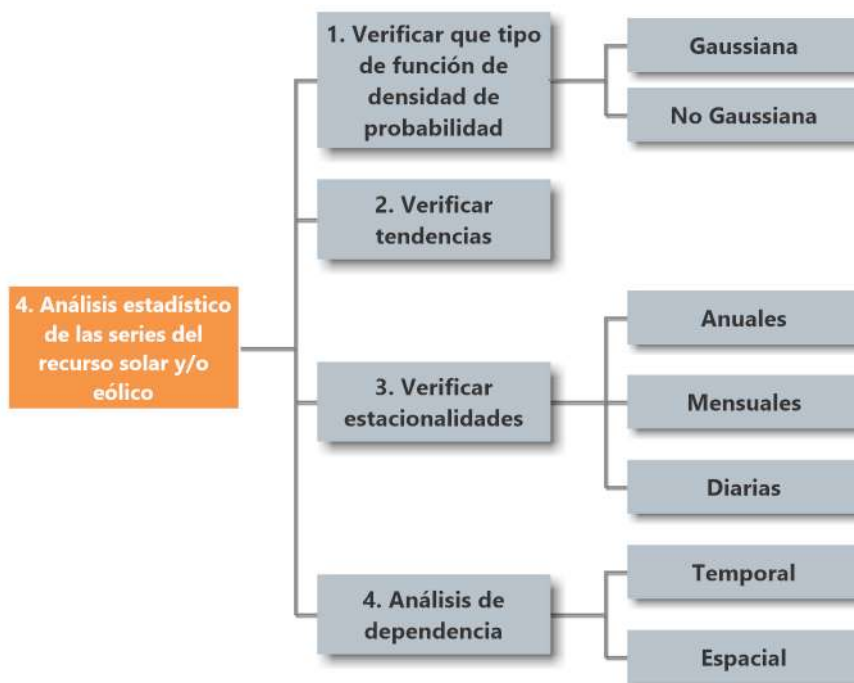


Figura 3.8. Mapa conceptual de la etapa 4: Análisis estadístico de las series del recurso solar y/o eólico

Verificar el tipo de función de densidad de probabilidad que tiene el recurso

Para verificar que las series de recursos se distribuyen de acuerdo a una normal se realiza en primer lugar una inspección visual de los histogramas asociados a cada serie, ello dirá si su función de densidad de probabilidad se acerca a una distribución normal. Si existen dudas se pueden realizar algunos test estadísticos más formales tales como Anderson-Darling y χ^2 . Sin embargo, la aplicación de estos test para verificar la normalidad de las funciones de distribución no es común debido a que es aceptado que las series de velocidad de viento, potencia eólica, radiación y potencia fotovoltaica tienen distintas funciones de distribución. Estas pruebas de hipótesis se aplican para verificar que las transformaciones escogidas sean adecuadas y efectivamente el nuevo proceso se distribuya de acuerdo a una normal [109].

Anderson-Darling y χ^2 sirven para verificar si la función de densidad de probabilidad que se estima es la adecuada. También pueden utilizarse para verificar si las series siguen una distribución Weibull, Beta, entre otras. Una prueba de hipótesis utilizada con frecuencia es la de Kolmorov-Smirnov, esta prueba no se ha considerado porque los valores de rechazo contemplan que los parámetros de la función de densidad no son estimados y para este caso se calculan los parámetros de la función de distribución. De allí que la prueba Anderson-Darling es más apropiada. Los detalles respecto al cálculo y aplicación pueden ser encontrados en [121]

No es necesario aplicar estas pruebas de hipótesis cuando el histograma claramente muestre asimetrías o no sea unimodal, este es el caso específico de la radiación solar, si los recursos que deben ser analizados incluyen radiación solar automáticamente las series serán no gaussianas. Estas características indicarían la no normalidad de los datos. Por otro lado, dado que la mayoría de los modelos son del tipo ARIMA(p,d,q) y VAR(p) con todas sus variantes y estos modelos necesitan que sean series normales y estacionarios, por lo tanto estas pruebas de hipótesis deben ser aplicadas a las series una vez que se hayan realizado los procesamientos previos a los datos [86].

Verificar tendencias

En la práctica las tendencias se pueden apreciar con mayor frecuencia en las series de potencia eólica o fotovoltaica debido a la instalación de nuevas centrales en los SEP[104]. Por lo tanto, cada vez que se tenga un análisis de una serie de potencia y exista un cambio en la capacidad instalada durante el tiempo, esta serie tendrá una tendencia.

Cuando no sea claro, se debe aplicar la prueba de hipótesis Dicker-Fuller aumentado. Ésta es una prueba de hipótesis que busca determinar si es necesario realizar una diferenciación en la serie en contra de la hipótesis alternativa, la cual consiste en que la serie de tiempo sea estacionaria o estacionaria con tendencia [122]. El rechazo indica que no hay evidencia estadística de la existencia de tendencia en la serie. Cuando la serie no posee estacionalidades se concluye además que es estacionaria [123]. Los detalles respecto al cálculo pueden ser encontrados en [122].

Variaciones estacionales

Esta subetapa es de gran importancia debido a que, como fue observado en el capítulo del estado del arte, algunas metodologías suponen que las series pueden ser descompuestas de acuerdo a la ecuación (2.8). Las componentes estacionales en series de radiación solar existen debido a la naturaleza del recurso y no es necesaria su verificación, mas sí su análisis. Por otro lado, en el caso de las series de velocidad de viento, es usual que se asuma la existencia de estacionalidades, pero éstas pueden o no ser significativas, lo que ha sido observado cuando se han utilizado cadenas de Markov [50]. En [50] se aplica una prueba de hipótesis para verificar si la matriz de transición de probabilidades no varía en el tiempo, los resultados muestran que no hay evidencia estadística suficiente para garantizar la existencia de estacionalidades marcadas.

Para verificar la existencia de estacionalidades, se utilizan gráficos de diagramas de cajas, medias horarias, mensuales o para el tiempo de muestreo del cual se disponga. En esta propuesta se realiza un énfasis mayor en el diagrama de cajas, puesto que es posible observar la distribución que tienen los datos en distintos horizontes de tiempos. Como ejemplo se muestra un diagrama de cajas en la Figura 3.9 para las variaciones horarias de velocidad de viento. Las cajas representan la variación entre cuartiles y la línea roja la mediana, de esta figura se desprende la existencia de cambios en la media y varianza para las distintas horas del día lo que implica que existe un patrón estacional horario. Esto puede ser verificado al aplicar las siguientes pruebas de hipótesis:

- Friedman
- Kruskal Wallis

En la prueba Friedman se desea verificar si las medias de las muestras asociadas a distintos grupos son idénticas. Se supone que las varianzas son iguales y que se distribuyen de acuerdo a una normal. Una prueba más concluyente es la de Kruskal Wallis, la cual no necesita que las muestras sigan alguna distribución conocida, sólo exige que tengan la misma forma [124]. En ambos casos se rechaza la existencia de una estacionalidad cuando los p-valores son menores a un nivel de significancia, valor que se ha fijado en 5%.

Ya que se desea saber si existe estacionalidad mensual, diaria, horaria o para tiempos de muestreo menores, si los diagramas de caja no son completamente concluyentes como se observa en la Figura 3.10, al aplicar las pruebas anteriores se puede tomar una decisión respecto a la existencia o no de estacionalidades. Las pruebas de hipótesis deben aplicarse a series sin tendencia [124]. El detalle de los cálculos de los estadísticos puede ser analizado en [124].

Las pruebas de hipótesis descritas en los párrafos anteriores sólo son necesarias si los métodos gráficos no son concluyentes. Sin embargo, en caso que las series tengan variaciones estacionales, estas pruebas de hipótesis son utilizadas para corroborar que los procesamientos previos a las series de tiempo han removido las variaciones estacionales.

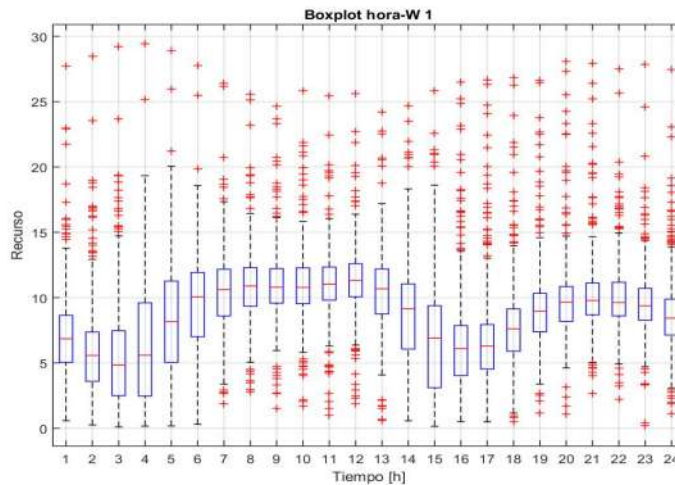


Figura 3.9. Diagrama de cajas de velocidad de viento en zona de Antofagasta (Datos obtenidos en [17])

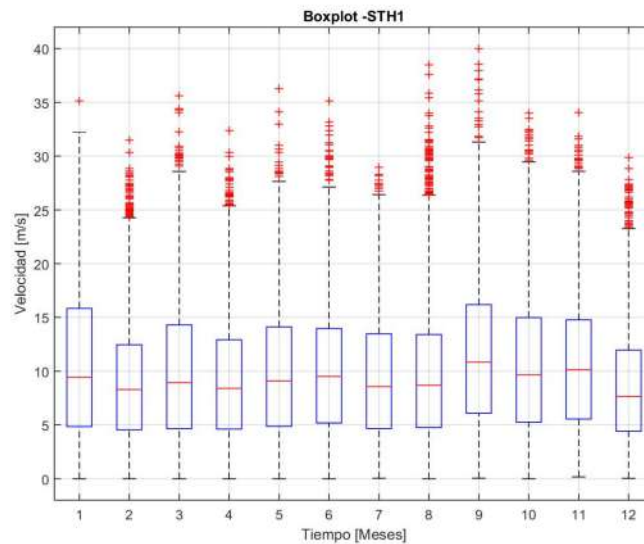


Figura 3.10. Diagrama de cajas de velocidad de viento mensual en Nueva Zelanda (Datos obtenidos en [125])

Análisis de dependencia temporal

La dependencia temporal puede ser observada a través de la función de autocorrelación simple y parcial (para mayor detalle ver Anexo 1). En la Figura 3.11 se muestra la función de autocorrelación simple y parcial de una serie de potencia eólica. Las líneas azules representan la banda de confianza. Cuando los coeficientes son mayores a esta banda, se puede concluir que existe evidencia estadística de que las autocorrelaciones son significativas [122].

Las bandas están definidas de acuerdo a la ecuación (3.1), en donde n es el número total de muestras u observaciones. Entonces existirá dependencia temporal si, al menos un coeficiente de correlación distinto al retardo cero, es mayor a los límites.

$$-\frac{1}{n} \pm \frac{2}{\sqrt{n}} \tag{3.1}$$

Si quedara alguna duda tras observar las gráficas de las funciones de autocorrelación simple y parcial, se puede aplicar la prueba de hipótesis Ljung-Box. Las dudas pueden surgir principalmente para coeficientes

de correlación que tomen valores ligeramente mayores a la banda de confianza, lo que llevaría a pensar que dicho coeficiente efectivamente no afecta considerablemente el modelo, cuando ello ocurre la aplicación de la prueba de hipótesis Ljung-Box se puede utilizar para confirmar la decisión de significancia del coeficiente.

Las hipótesis nula y alternativa son las siguientes:

Hipótesis nula $H_0: \rho_1 = \dots = \rho_h = 0$

Hipótesis alternativa $H_a: \rho_i \neq 0$ para algún i que satisface $1 < i \leq h$

Por lo tanto, si existe evidencia estadística para rechazar la hipótesis nula, entonces se concluye que existe dependencia temporal. Una observación debe ser realizada respecto a la dependencia temporal: los recursos eólicos y solares tienen correlación temporal, por ende el graficar estas funciones debiera ser suficiente para verificar la existencia de dicha correlación.

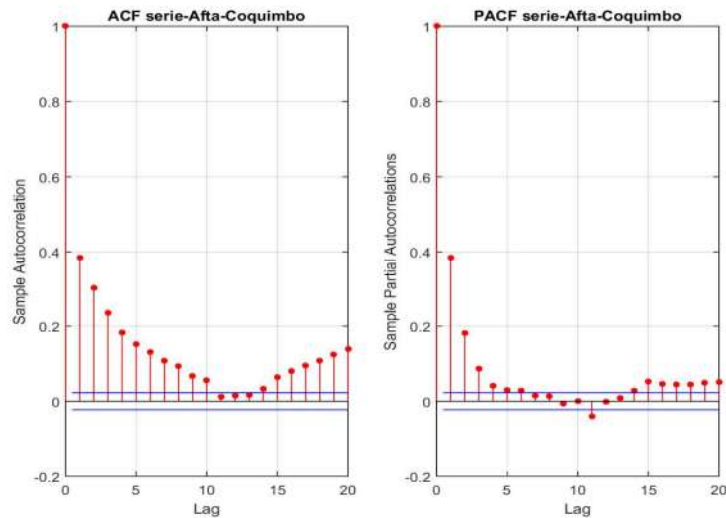


Figura 3.11. Ejemplo de función de autocorrelación de serie de potencia eólica agregada en las zonas de Antofagasta a Coquimbo.

Análisis de dependencia espacial

La existencia de dependencia espacial es un poco más compleja de verificar. Es posible calcular la matriz de correlación, sin embargo el coeficiente de correlación es solo el segundo momento de la función de densidad de probabilidades en el retardo cero. Para poder definir la existencia de correlación espacial se propone utilizar la función de autocorrelación cruzada como se muestra en la Figura 3.12. Al igual que la función de autocorrelación simple esta contiene dos bandas de significancia las cuales sirven para definir la existencia de correlación espacial.

Cuando existan dudas o la cantidad de series aumente es aconsejable utilizar la prueba de hipótesis Ljung-Box en su versión multivariable [126]. Las hipótesis son idénticas a su versión univariable, el cambio consiste que, en vez de utilizar los coeficientes de autocorrelación, se utilizan las matrices de autocorrelación cruzada. Existirá correlación entre las variables cuando los p-valores sean menores a un nivel de significancia, se utilizará un valor común del 5%, como se muestra en la Figura 3.12 (ver Anexo 1 para detalles para el cálculo y estimación de la función de autocorrelación cruzada y en [126] el detalle de la aplicación de la prueba de hipótesis).

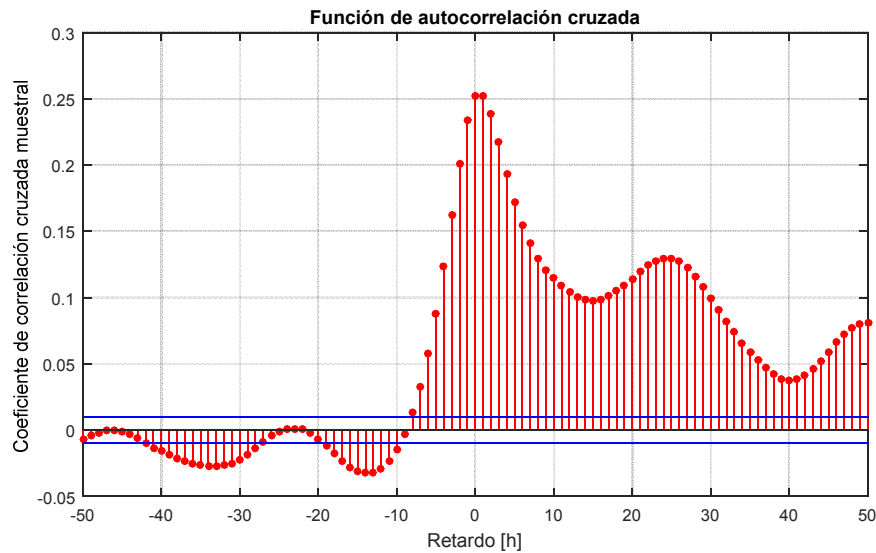


Figura 3.12. Función de autocorrelación cruzada de dos series de Nueva Zelanda (datos obtenidos de [125])

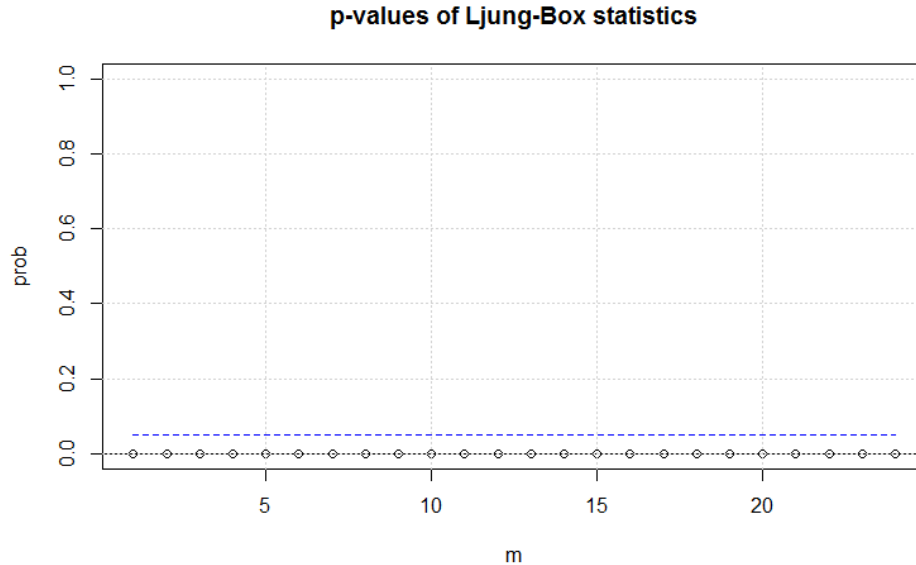


Figura 3.13. Resultado de prueba de hipótesis Ljung-Box para dos series correlacionadas espacialmente

Estacionariedad

Es posible concluir si las series son estacionarias en función del análisis estadístico previo. Si la serie no contiene estacionalidades ni tendencias y es gaussiana entonces será estacionaria. Si alguna de estas características es positiva, entonces inmediatamente no será estacionaria.

Diagrama de flujo del análisis estadístico

El diagrama mostrado en la Figura 3.14 muestra el proceso que se debe realizar en la etapa de análisis estadístico. Esta etapa de gran importancia porque permite realizar un filtro en la cantidad de modelos a considerar. Además, las pruebas estadísticas mostradas en esta etapa se vuelven a utilizar para evaluar los supuestos de todos los modelos del tipo ARMA y VAR.

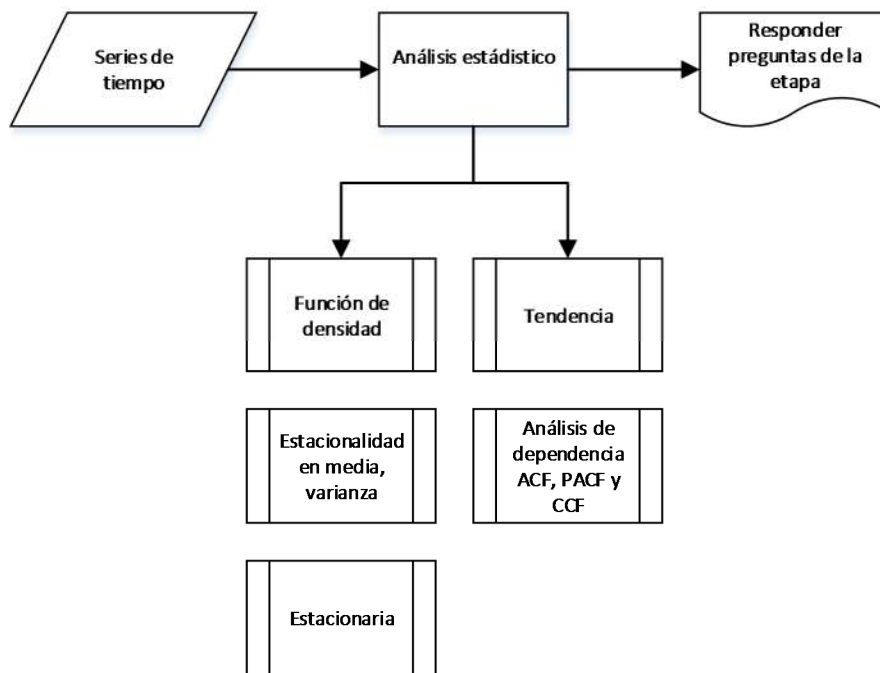


Figura 3.14. Diagrama de flujo de análisis estadístico

Cuestionario

Las preguntas que se deben responder en función del análisis estadístico se detallan a continuación, para cada pregunta se considera válida una alternativa.

1. ¿Qué tipo de función de densidad de probabilidades tienen los recursos?

Las respuestas válidas a esta pregunta son:

- a) Función Gaussiana
- b) No Gaussiana

2. ¿Tiene alguna de las series tendencia?

Las respuestas válidas a esta pregunta son:

- a) Sí
- b) No

3. ¿Tiene alguna de las series variaciones estacionales?

Las respuestas a esta pregunta pueden ser:

- a) Sí
- b) No

Se considera válida una alternativa

4. ¿Existe dependencia temporal?

Las respuestas a esta pregunta pueden ser

- a) Sí
- b) No

5. ¿Existe dependencia espacial?

Las respuestas a esta pregunta pueden ser

- c) Sí
- d) No

6. ¿Son todas las series estacionarias?

Las respuestas a esta pregunta pueden ser

- a) Sí
- b) No

En la Figura 3.15 se muestra la relación de las preguntas de esta etapa con el resto de las etapas de la metodología propuesta.

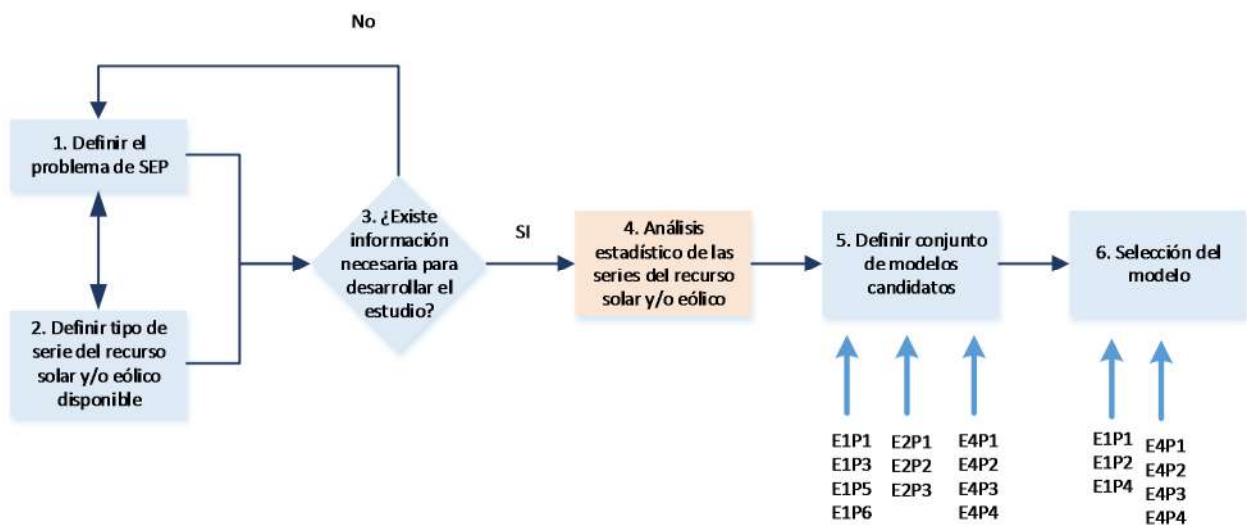


Figura 3.15. Relación entre preguntas de la etapa 4 con el resto de las etapas de la metodología propuesta

3.2.5 Definir conjunto de modelos candidatos

El esquema general de la etapa 4 se muestra en la Figura 3.16. En él se observan en términos generales los modelos según dependencia y recurso. El objetivo de esta etapa es proponer un conjunto de modelos que ya han sido desarrollados y presentados en la literatura e incluye variantes que pueden ser aplicadas a los modelos ARMA y VAR. Para cumplir con este objetivo se ha confeccionado, para los distintos modelos, una base de datos detallada en las tablas del Anexo 4. Para lograr acotar el conjunto de modelos se utilizan las respuestas a las preguntas de las etapas 1,2 y 4 y las series de tiempo obtenidas en la etapa 2. Teniendo el tipo de recurso e información estadística se realiza una búsqueda en la base de datos de tal forma que coincida el problema actual con los modelos, obteniendo así el o los candidatos.

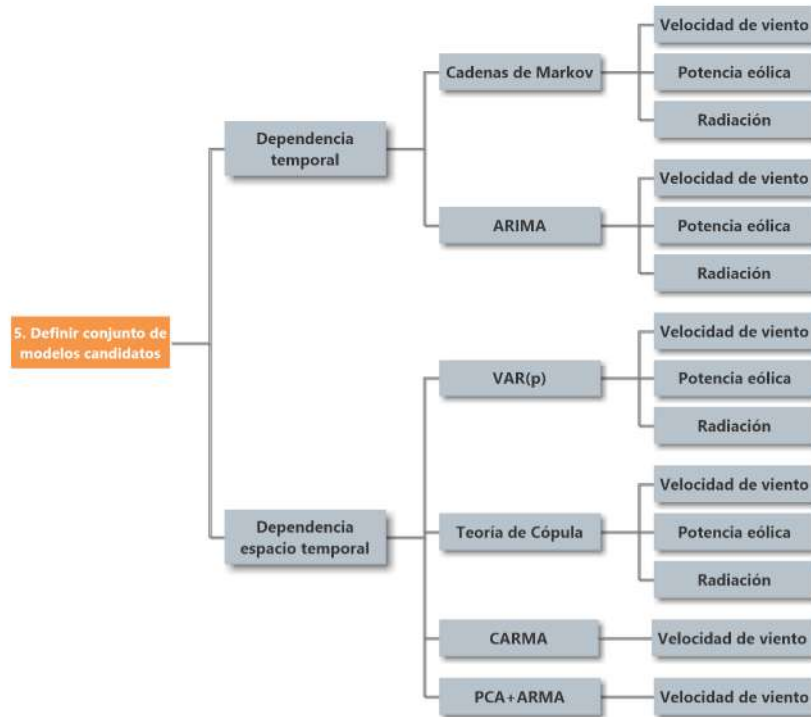


Figura 3.16. Mapa conceptual etapa 4: Definir conjunto de modelos candidatos.

El diagrama de flujo que muestra el proceso a seguir en esta etapa para acotar los modelos candidatos se encuentra en la Figura 3.17, y se detalla a continuación:

1. Acotar la selección de modelos en función de la existencia de correlación espacial. Si la respuesta a la pregunta respecto a la correlación espacial es positiva, entonces la búsqueda será sólo entre los modelos que consideren esta característica. De igual forma, si existe un conjunto de series que no posee correlación espacial, entonces para estas series se buscará sólo entre los modelos que consideren correlación temporal.
2. Entre los modelos acotados en el paso 1 buscar todos aquellos modelos propuestos que hayan sido probados en las series de tiempo de recurso (velocidad, radiación potencia eólica y solar) que son de interés para el problema de SEP que se desea resolver. En esta metodología sólo se consideran los modelos que han sido analizados en la literatura para los recursos y no se consideran las extensiones que hacen los autores a otros recursos sin mostrar resultados. Por ejemplo, en [110] se indica que el modelo sería posible aplicarlo a series de radiación, pero no comenta las variaciones que pueden ser realizadas ni existen resultados utilizando dicha técnica, por ello para dicho modelo sólo se considera que es posible utilizarlo para series de velocidad de viento.
3. Si existirán plantas nuevas de cualquier índole en el horizonte de evaluación, entonces utilizar sólo aquellos modelos que consideran el recurso primario de velocidad o radiación. De lo contrario, escoger series de potencia ya que se evita error de conversión sobre todo para la energía eólica [53]
4. Realizado el paso 2, utilizar aquellas metodologías que consideran las características de las series de tiempo del recurso disponible (tendencia y estacionalidad). Las bases de datos de modelos contienen información referente a si las metodologías de los artículos consideran efectos estacionales, en caso que existan. Cuando se consideren modelos espacio temporales, bastará con que una serie tenga tendencia o estacionalidad para seleccionar el modelo que considera dichas características

5. Si las series no son Gaussianas, seguir acotando a aquellos modelos que consideren procesamiento previo o no necesiten dicha consideración para poder ser aplicados.
6. Si existe correlación espacial en los datos, verificar si todas las series que se encuentran correlacionadas tienen igual longitud de datos. Los modelos VAR(p), PCA+ARMA(p,q) y teoría de Cópula necesitan que los datos se encuentren sincronizados. En cambio, los modelos CARMA(p,q) con sus variantes, pueden ser aplicados cuando la longitud de las series de tiempo son distintas, pero tienen un desempeño menor en la caracterización de la correlación espacial [7]. Siempre será mejor o ideal disponer de datos sincronizados.

De acuerdo a lo observado en la Figura 3.16, cuando se está en presencia de energía solar y eólica y existe correlación espacial, la selección es directa debido a que han sido reportados dos modelos para generar series sintéticas: el trabajo presentado en [86] que utiliza un modelo VAR(p) y en [109] que utiliza teoría de Cópula. En cambio, cuando se está en presencia de múltiples sitios de energía eólica, la cantidad de modelos aumenta. Para efecto de este trabajo se han implementados modelos MCMC, ARMA(p,q), CARMA(p,q) (dos variantes), VAR(p), PCA+ARMA(p,q) y función de Cópula Gaussiana. No obstante, se ha dado mayor énfasis al modelo VAR(p) por ser el más utilizado en la generación de series sintéticas y se ha observado que a través de procesamientos previos es posible aplicar y tener buenos resultados [7], [34], [86], [100]–[105], [127]–[129]. Los algoritmos y teoría asociados a ellos se encuentran en el Anexo 2.

El procedimiento descrito puede ser seguido fácilmente utilizando la información obtenida en las etapas 1, 2 y 4 como se observa en la Figura 3.15. En la Figura 3.17 se muestra un diagrama del proceso, en donde se puede observar en qué etapas se utilizan las preguntas de las etapas anteriores. Una vez definido qué series se encuentran correlacionadas y cuáles no, es posible aplicar el proceso para cada conjunto de series.

3.2.6 Selección del modelo

En esta etapa se busca determinar un modelo adecuado para ser utilizado en el problema de planificación de la inversión u operación. Dicha búsqueda se realiza entre los modelos candidatos entregados por la etapa anterior. Es importante destacar que esta es una tarea compleja, debido a que toda decisión depende de las características específicas de las series de tiempo que se están analizando, el problema de SEP y el horizonte de evaluación, lo que hace que la decisión pueda considerar diversas alternativas de modelos, pudiendo ser válidas cada una de ellas. A pesar de ello, para poder cumplir con el objetivo de esta etapa se propone un esquema en la Figura 3.18, siguiendo dicho orden es posible determinar un modelo adecuado.

En la literatura no existe un indicador universal, ni se utilizan los mismos estadísticos para analizar la bondad de cada modelo. No obstante, considerando que el objetivo de las series sintéticas consiste en generar series de tiempo que mantengan las principales propiedades estadísticas de las series de tiempo originales, se utilizarán los estadísticos más importantes, los cuales son:

- Promedio aritmético
- Varianza
- Desviación estándar
- Cuantiles 10%, 25%, 50%, 75% y 90%

A los estadísticos se agrega las funciones: (el cálculo de estas funciones se encuentra en el Anexo 1):

- Función de densidad de probabilidad. Para esta función se utiliza el histograma de frecuencias.
- Función de autocorrelación simple
- Función de autocorrelación parcial
- Matriz de autocorrelación cruzada

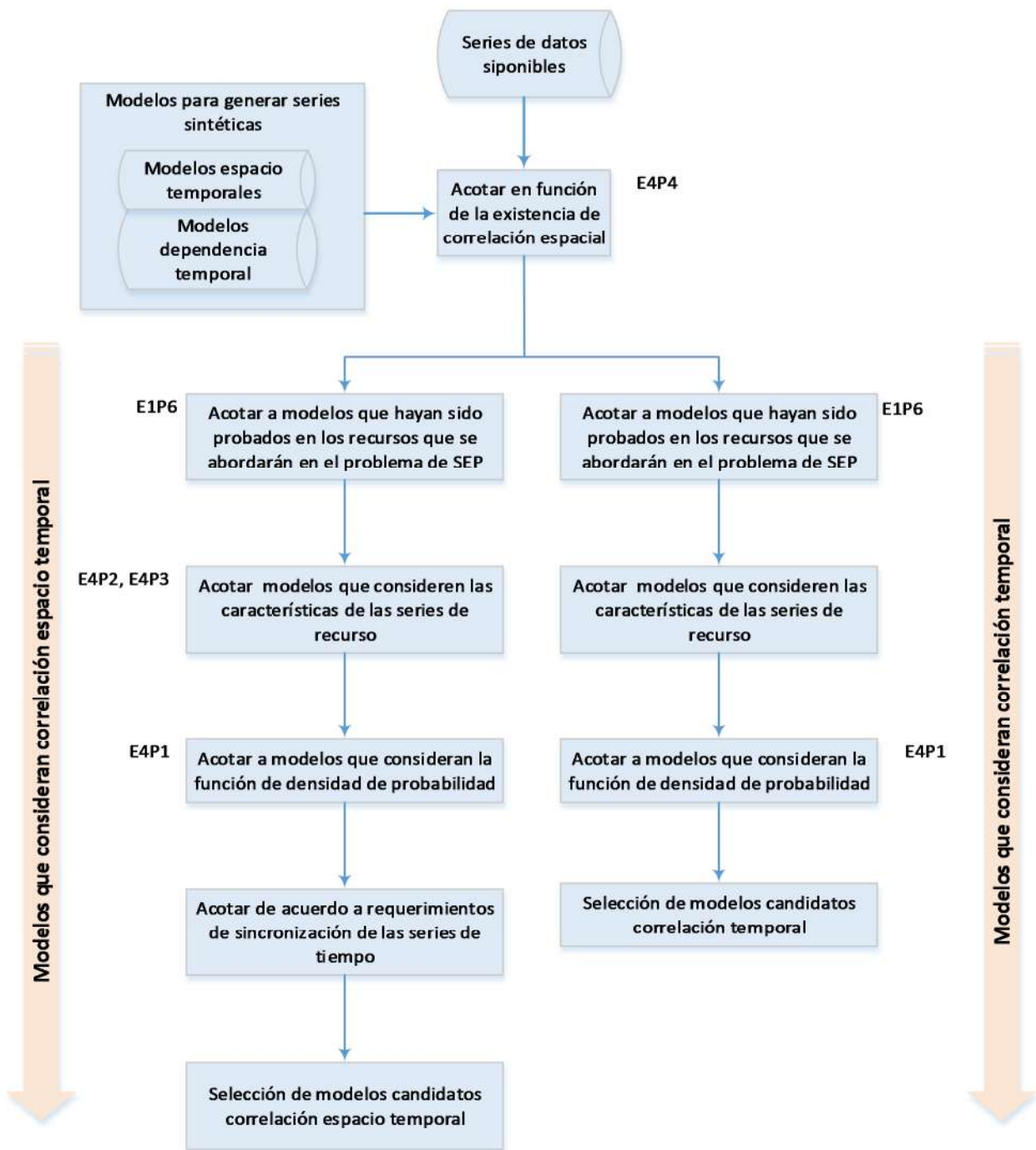


Figura 3.17. Diagrama de proceso para seleccionar modelos candidatos

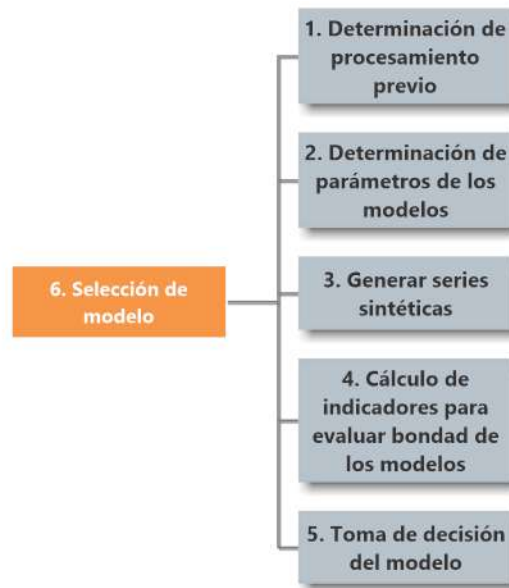


Figura 3.18. Mapa conceptual etapa 6: Selección de modelo

Con estos estadísticos y funciones es posible verificar qué tan bien el modelo caracteriza la serie real. Se ha observado en la literatura que, en general, cuando se analizan los estadísticos de toda la serie sintética y se comparan con los de la serie de tiempo real, éstos manifiestan un error bastante pequeño lo que hace difícil en base a ellos tomar una decisión respecto al mejor modelo [50][86], [102], [104]. Por otro lado, las diferencias se presentan principalmente en la función de autocorrelación y función de densidad de probabilidad, funciones que dan a conocer si la serie sintética representa adecuadamente la persistencia que existe en los recursos.

Además de realizar un análisis estadístico sobre la totalidad de las muestras se reconoce la necesidad de analizar muestras más pequeñas. Por ejemplo, para un caso en que las series de tiempo son horarias se deberá calcular cada estadístico para $t=1, \dots, 24$, tanto en la serie real como sintética y compararlos. Con ello, se obtendrá información que permita diferenciar claramente las bondades de cada modelo para los tiempos de simulación específicos.

En problemas de planificación de la operación de mediano, largo y muy largo plazo, se deben calcular los estadísticos para distintos horizontes de tiempo. Por ejemplo, estadísticos horarios y mensuales, cuando se tienen datos muestreados cada hora y el horizonte es mayor a un mes. En algunos trabajos se han observado la media y la desviación estándar, lo que equivale al primer y segundo momento de la función de densidad de probabilidades. Sin embargo estos momentos no representan el comportamiento completo de sus funciones de densidad de probabilidades, por ello en este trabajo se considera apropiado agregar los cuantiles, los cuales entregan información de la función de distribución de probabilidades.

La pregunta que surge en este momento es ¿Cómo se pueden comparar modelos? Al ser distintos los parámetros que deben ser calculados y además se observan diversas gráficas, es una tarea compleja realizar una comparación y escoger el mejor modelo. Puede darse que todos los modelos candidatos sean adecuados y su aplicación dependerá de las competencias del equipo que debe realizar el estudio.

En la literatura, a diferencia de los modelos de pronósticos, no existe un consenso de indicadores que puedan ser considerados en forma sistemática para realizar comparación entre los modelos para generar series sintéticas. En esta metodología se utilizan definiciones propuestas para modelos de pronósticos y se aplican sobre los estadísticos definidos en los apartados anteriores. Se ha considerado razonable utilizar medidas de error el cual se define:

$$e_k = x_k - \hat{x}_k \quad (3.2)$$

Donde x_k será el valor estadístico obtenido de los datos históricos y \hat{x}_k es el valor obtenido a través de la serie sintética. En ocasiones es más fácil considerar los errores normalizados, lo que implica que la ecuación (3.2) se transforma en:

$$e_{k0/1} = \frac{x_k - \hat{x}_k}{x_{\max}} \quad (3.3)$$

Considerando la definición de error se definen los siguientes indicadores [31]:

- Mean square error (MSE):

$$MSE = \frac{1}{N} \sum_{k=1}^N e_k^2 \quad (3.4)$$

- Root mean square error (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{k=1}^N e_k^2} \quad (3.5)$$

- Root mean squared relative error (RMSRE):

$$RMSRE = \frac{RMSE}{\sqrt{\frac{1}{N} \sum_{k=1}^N x_k^2}} \quad (3.6)$$

Con la descripción de los estadísticos, funciones e indicadores se explicará cada subetapa de esta etapa, hasta llegar a la selección.

Determinación de procesamiento previo

En esta subetapa se desea definir el tipo de transformaciones que pueden ser utilizadas cuando los modelos candidatos lo necesiten. El esquema general se muestra en la Figura 3.19. Las opciones que se han agregado que permiten remover tendencias, variaciones estacionales y transformar series no Gaussianas a Gaussianas son las que se han propuesto con mayor frecuencia en la literatura especializada [9][34] [86]. Esta subetapa solo se debe aplicar cuando las series posean alguna de estas características, de lo contrario se pasa a la siguiente subsección. Las técnicas que se han especificado en el esquema son aquellas que han sido utilizadas frecuentemente y para poder aplicarlas adecuadamente se hace necesaria la información obtenida en el análisis estadístico de la etapa 4.

Remover las tendencias y estacionalidades son exigencias de los modelos ARMA y VAR para convertir las series de tiempo a series estacionarias. Convertir las series de tiempo a Gaussianas se realiza por dos razones importantes: en primer lugar para que el ajuste y sintetización permita caracterizar adecuadamente la función de densidad de probabilidades y en segundo lugar, si el proceso es gaussiano, entonces el proceso será estrictamente estacionario, lo que garantiza la convergencia del proceso [126].

Las tendencias en las series sin correlación espacial han sido removidas a través de una diferenciación como muestra la ecuación (3.7):

$$Y_t = X_{t-i} = B^i X_t \quad (3.7)$$

Cuando se trata de un proceso multivariable, independiente que se considera diagonalizada la matriz, no se han utilizado procesos cointegrados. El único trabajo que ha presentado una metodología para tratar las series con dichas características es [104]. En este trabajo se aplica un modelo VAR sobre series de potencia eólica, las cuales manifiestan tendencia debido al aumento de la capacidad instalada. Para abordar el problema se define una nueva variable que corresponde a la proporción de energía eólica y se define como:

$$Y_t = \frac{X_t}{P_t} \quad (3.8)$$

Donde X_t y P_t corresponde a la serie de tiempo de potencia eólica y a la variación en la potencia instalada. En el artículo se utiliza una aproximación lineal para estimar P_t , teniendo la capacidad instalada al iniciar y finalizar las mediciones. Existen más técnicas que podrían ser utilizadas para modelar las tendencias, pero en la literatura sólo se ha realizado dicha aproximación hasta la fecha.

La estacionalidad puede ser removida mediante distintos procesamientos de datos, los cuales se describieron en el estado del arte y para simplificar la explicación se vuelven a mostrar en las ecuaciones (3.9)-(3.11). El objetivo de estos procesamientos es obtener series estacionarias. En algunos casos, tras remover la estacionalidad, el proceso también termina Gaussiano, pero no es el caso general.

Si existen variaciones en la media, pero no en la varianza se puede utilizar (3.9). Si además la varianza varía a través del tiempo se deberá aplicar (3.10) en donde el conjunto S hace referencia a las estaciones consideradas. Como es usual utilizar series con tiempo de resolución de 1 hora, S puede ser el conjunto de horas por mes, así $\mu_{tk} = 24 * 12 = 288$ variables estacionales.

Otra opción para remover la estacionalidad consiste en utilizar la transformada integral para cada variable estacional. En [86] se utiliza en una serie de mediciones horarias. Para caracterizar las variaciones horarias se utilizan 24 transformaciones, lo que implica una función de distribución de probabilidad estimada por hora. Tiene excelente resultados, sin embargo, si las series tuvieran variaciones estacionales mensuales, entonces se necesitarían 288 funciones de distribución y ello implica que se debe disponer de muchos años de información para que dicha estimación sea buena [127]. No obstante lo anterior, si la longitud es cercana a un año, es posible utilizar las ecuaciones (3.9) y (3.10) para remover la estacionalidad mensual y diaria; y para estacionalidades horarias o menores utilizar transformada integral. Esto no ha sido aplicado en artículos de investigación pero en los siguientes capítulos de esta tesis se mostrará que se obtienen excelentes resultados.

$$Y_t = X_t - X_{at} - X_{dt} \quad (3.9)$$

$$X'_t = \frac{X_t - u_{tk}}{\sigma_{tk}}, \quad \forall k \in \{S\} \quad (3.10)$$

$$Y_t = F_{invN,k}^{-1}(F(X_t)), \quad \forall k \in \{S\} \quad (3.11)$$

La transformada integral necesita de la estimación de la función de distribución acumulada, la cual puede ser estimada utilizando:

- Función de distribución empírica: es de fácil aplicación, pero la desventaja es que en el proceso de simulación cuando las series sintéticas son generadas, no es posible obtener valores fuera del rango medible
- Función de distribución estimada a través de Kernel-Gaussiano: Corresponde a un ajuste suavizado sobre la función de distribución empírica. Requiere mayor procesamiento y el ajuste depende del ancho de banda que se utilice.
- Función de distribución empírica ajustada a funciones conocidas: Las series de velocidad de viento y radiación pueden ser ajustada a funciones de densidad de probabilidad Weibull y beta. Utilizar

estas funciones solo es posible cuando se logra verificar que los datos se ajustan específicamente a estas curvas

- Función de distribución usando Kernel más Pareto Tail: Entre los cuantiles 10 y 75% se ajusta una curva utilizando Kernel y en los extremos la función de distribución Pareto Tail. Es más compleja la implementación y no se aprecian mejores resultados en la estimación de los estadísticos en comparación al uso exclusivo de Kernel.

Para transformar las series de tiempo a Gaussianas se puede utilizar la transformada integral. Ésta puede ser utilizada después que se han removido la tendencia y estacionalidad. También se puede aplicar la transformación Box-Cox. Esta última no garantiza que el proceso sea gaussiano, a diferencia de la transformada integral. No obstante, la transformada integral tiene como requisito que la función de distribución acumulada sea estrictamente creciente [130].

El procedimiento para escoger transformada para datos previos se describe en la siguiente secuencia:

1. Remover tendencias. Si se han escogido modelos multivariados utilizar la ecuación (3.8) de lo contrario utilizar diferenciación para los modelos univariados
2. Para cada serie X_{it} con $i = 1, \dots, K$, siendo K el número de sitios, remover estacionalidad utilizando ecuaciones (4.8)-(4.9)
3. Aplicar prueba de hipótesis Kruskal Wallis. Si pasan esta prueba de hipótesis entonces se ha removido la estacionalidad. De lo contrario aplicar la ecuación (4.10). Si no se dispone de los datos suficientes entonces utilizar ecuación (4.8) para las variaciones mensuales y (4.10) para las variaciones horarias.
4. Aplicar prueba de hipótesis Anderson-Darling para verificar normalidad. Si los datos no son Gaussianos después de haber removido estacionalidad, utilizar transformada integral y volver a verificar si los datos son Gaussianos.

La aplicación de la transformada integral puede ser vista con mayor detalle en el Anexo 1 de este trabajo, en él se muestra ejemplos gráficos y matemáticos de su aplicación.

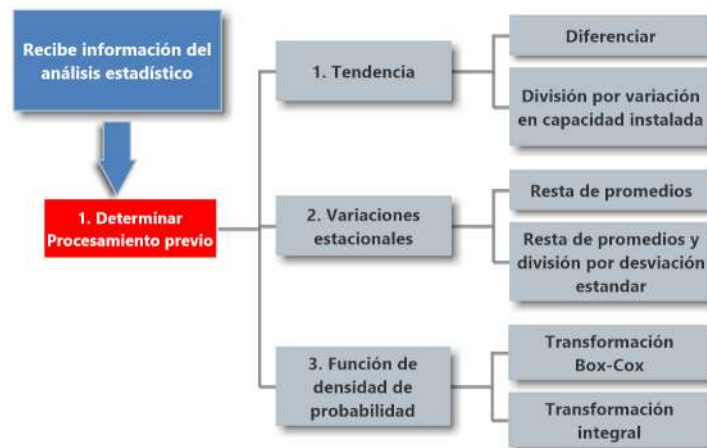


Figura 3.19. Subetapa 1: determinación de procesamiento previo

Determinación de parámetros de los modelos

En esta subetapa se calculan los parámetros de los modelos candidatos. En la Figura 3.20 se muestra el esquema general de esta subetapa. En ella solo se muestran los modelos que son considerados en este trabajo de tesis. Sin embargo, queda abierta la posibilidad de agregar más metodologías de generación de series sintéticas.

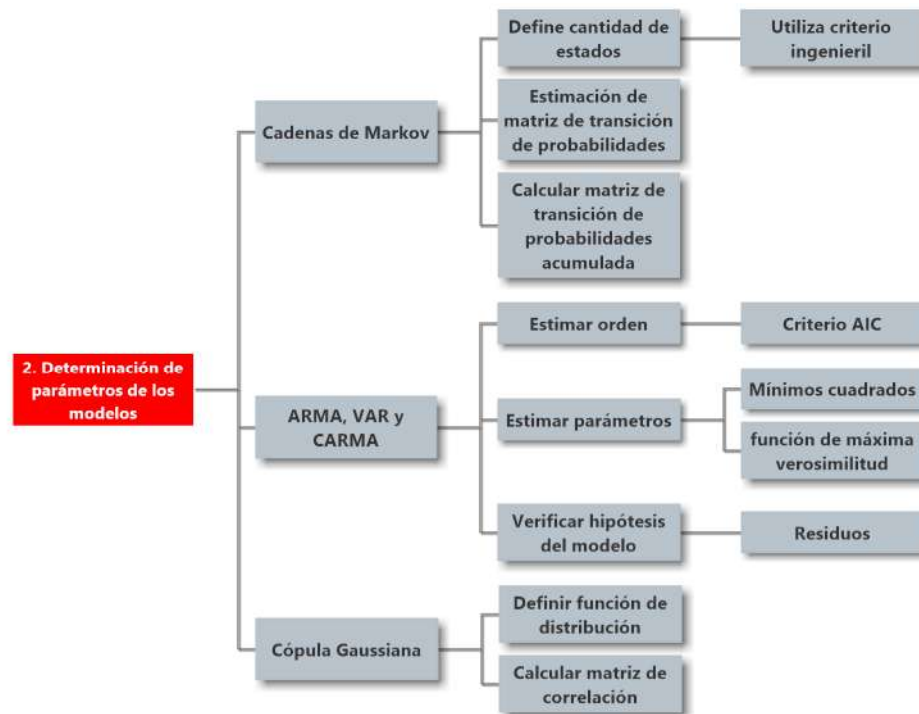


Figura 3.20. Subetapa 2: determinación de parámetros de los modelos

Para las cadenas de Markov se deben realizar los siguientes pasos:

1. Definir el conjunto de espacios de estados.
2. Obtener la matriz de transición de probabilidades.

Definir el conjunto de estados es una tarea que se realiza frecuentemente a través de criterio ingenieril. En la literatura se ha utilizado distinta cantidad de estados para representar los datos y se ha observado que cuando aumenta la cantidad de estados mejora la bondad del modelo pero aumenta la complejidad.

Para los modelos ARMA(p,q) y VAR(p) se deben realizar los siguientes pasos para obtener el orden y parámetros del modelo:

1. Utilizar criterio AIC para tener la primera estimación de los valores que tomará p y q (recordar que para los modelos VAR sólo existen parámetros p). Las ecuaciones para el cálculo se presentan en el anexo 1.
2. Estimar los parámetros utilizando método de mínimos cuadrados para los casos en que no existe componente media móvil. Para los demás utilizar la función de máxima verosimilitud.
3. El criterio AIC indica el compromiso que existe entre la complejidad del modelo y bondad del ajuste pero no indica si las hipótesis respecto a los residuos de los modelos son adecuadas. Es por esto que se debe verificar que los residuos sean normales y no correlacionados. Para ello se aplican pruebas de hipótesis de autocorrelación y autocorrelación cruzada, tal y como fueron descritas en la etapa de análisis estadístico. En caso que no se cumplan los supuestos, se debe ir incrementando el orden del modelo hasta que dichos supuestos se logren cumplir.

La estimación de los parámetros se puede realizar con cualquier programa estadístico y la formulación puede ser encontrada en diferentes libros referentes a series temporales. En caso de necesitar mayor información respecto a la estimación se aconseja revisar las referencias [122], [126], [131].

Los pasos descritos se relacionan con el método Box-Jenkins para estimar el orden de los modelos ARMA y VAR pero se han agregado las pruebas específicas que han sido aplicadas para asegurar que los modelos sean estimados adecuadamente. En el anexo 3 se muestra un ejemplo de la estimación del orden p para un modelo VAR utilizando el criterio AIC en su versión multivariable.

Un aspecto importante que se comentará a continuación es referente a los modelo VAR(p). Ya hemos dicho que estos modelos han dado excelentes resultados en caracterizar los recursos y se han propuesto distintas metodologías de procesamientos previos y transformaciones para ser posible aplicar esta estructura matemática. Dada la relevancia de este modelo en la generación de series sintéticas de recursos correlacionados espacio temporalmente, se ha desarrollado un diagrama de flujo que permite sistematizar el ajuste del modelo VAR(p), este se muestra en la Figura 3.21.

Antes de explicar el diagrama de flujo se recordará que modelo tipo VAR(p) tiene la siguiente estructura:

$$\vec{Y}_t = \sum_{i=1}^p A_i \vec{Y}_{t-i} + \vec{\varepsilon}_t \quad (3.12)$$

Donde A_i corresponde a la matriz de coeficientes modelo VAR (p), de orden $k \times k$, siendo k el número de series de tiempo y $\vec{\varepsilon}_t$ es un vector ruido blanco, el cual se distribuye en forma normal con matriz de covarianza no singular Σ . Para poder aplicar este modelo el vector \vec{Y}_t debe ser estacionario y Gaussiano.

En el diagrama de la Figura 3.21 se muestra el diagrama de flujo con el proceso para lograr estimar adecuadamente los parámetros de un modelo VAR a partir de datos disponibles de los recursos. Como se puede observar existe un proceso en donde se aplican transformaciones a las series de recursos, el propósito de esta etapa es transformar el proceso que podría ser no Gaussiano y no estacionario a uno que sí posee dichas características. El diagrama muestra que una vez realizada la transformación se deben aplicar las pruebas estadísticas correspondientes con el propósito de verificar si han sido removidas las estacionalidades y si son gaussianas las series de tiempo. Las pruebas de hipótesis que se deben aplicar son:

- Kruskal Wallis para verificar que las estacionalidades han sido removidas
- Anderson Darling para verificar normalidad
- Dicker-Fuller para verificar que la tendencia ha sido removida

Las pruebas anteriores son las mismas que fueron mencionadas en el análisis estadístico. Una vez aplicada las transformaciones se procede a estimar el orden del modelo a través de la utilización del criterio AIC (el procedimiento y ejemplo de cálculo se encuentra en [126]). Este criterio es un indicador entre la bondad del modelo y la complejidad, aunque referente a la bondad no es muy definitivo, por lo tanto se deben evaluar los residuos para verificar que no exista autocorrelación simple y cruzada. Para verificar esta condición se deben observar las funciones de autocorrelación simple y cruzada. Si después de observar las funciones de autocorrelación existen dudas respecto a éstas, se deberá aplicar la prueba de hipótesis Ljung-Box en sus formas univariable y multivariable. Como se muestra en el diagrama de la Figura 3.21, si alguna de las pruebas es rechazada se deberá aumentar el orden p del modelo hasta que ambas pruebas no rechacen sus hipótesis nulas.

Una vez que el orden del modelo ha sido determinado se debe verificar si los residuos son gaussianos. Para ello se vuelven a aplicar las pruebas de hipótesis de normalidad descritas en la sección de análisis estadístico de este capítulo. Si los residuos no se distribuyen en forma gaussiana, sería más apropiado no utilizar la factorización de Cholesky para realizar el proceso de muestreo, debido a que la serie sintética es generada a través de dicho proceso y es importante que quede bien representado [104]. Para ello en el diagrama de la Figura 3.21 se agrega un proceso en el cual, en caso en que los errores no sean gaussianos, será más apropiado utilizar teoría de cópula para generar el muestreo. En la práctica se ha observado que la función de distribución t-student puede ser utilizada en estos casos. Para realizar el muestreo se utiliza una Cópula Gaussiana (para mayor detalle ver anexo 1 y [127]). Es posible utilizar otras familias de Cópula, ya que se ha mostrado en la práctica que puede entregar buenos resultados [104], [105], [127]. El uso de esta

metodología para estimar los parámetros de un modelo VAR busca disminuir el ajuste de todos los parámetros de los modelos.

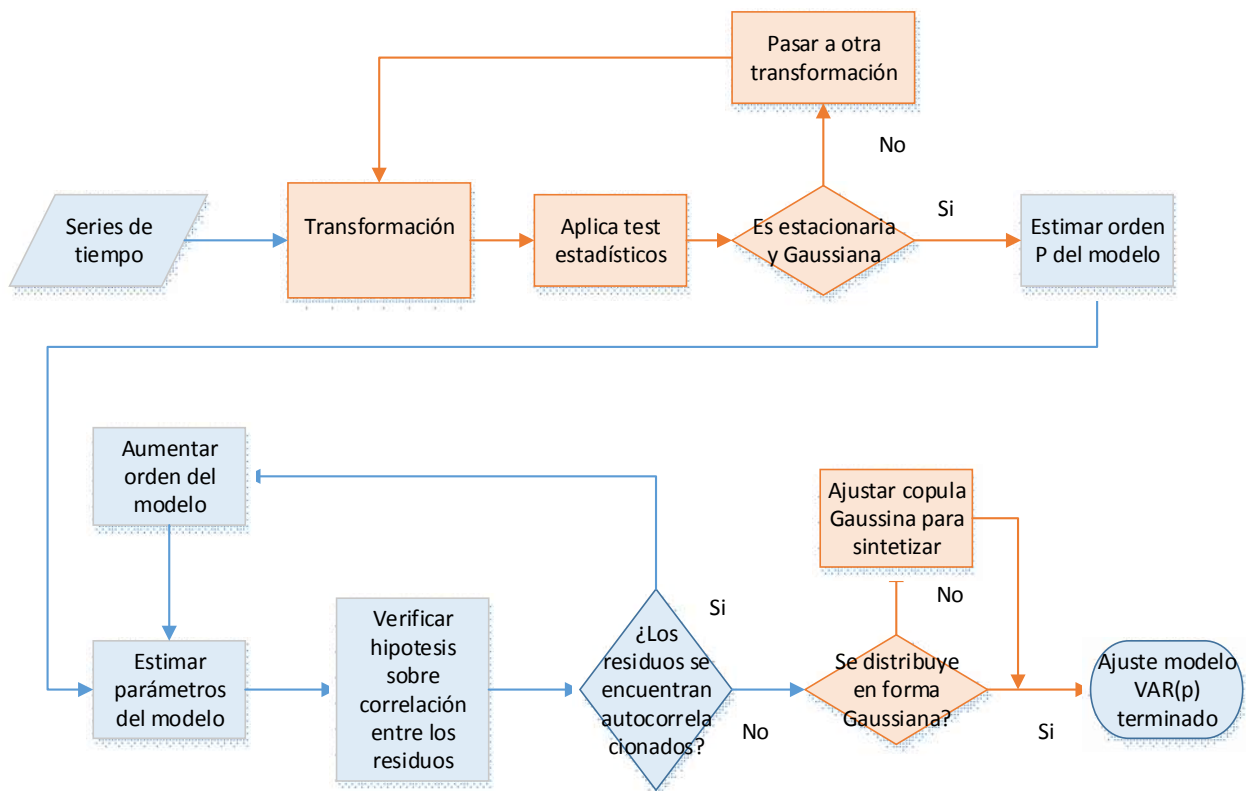


Figura 3.21. Diagrama de flujo para selección y ajuste modelos VAR(p)

Generación de series sintéticas

El esquema general de esta subetapa se muestra en la Figura 3.22. Recibe como datos de entrada los modelos ajustados de la subsección anterior y tiene como objetivo generar series sintéticas que tengan una longitud igual a la serie real. Se busca que tenga igual longitud porque cada realización de la serie sintética debe representar la estadística del proceso original y para comparar esta característica es apropiado mantener la longitud de la serie [132]. Como se muestra el esquema de la Figura 3.22 la generación de la serie sintética es propio a cada modelo. Los algoritmos que han sido utilizados para generar series sintéticas en este trabajo se muestran en el anexo 2 para los modelos temporales y en el anexo 3 para los modelos espacio temporal.

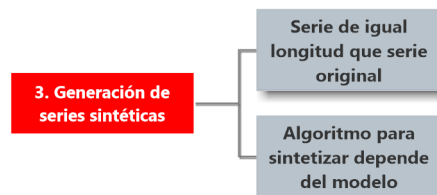


Figura 3.22. Subetapa 3: generación de series sintéticas

Cálculo de indicadores para evaluar la bondad de los modelos y comparar

Esta subetapa tiene como objetivo calcular los estadísticos e indicadores que serán utilizados para tomar la decisión de qué modelo utilizar cuando exista más de un modelo propuesto en la etapa 4. Los datos de entrada a esta sección son: la serie de recursos a las cuales se le ajustaron los modelos y las series sintéticas generadas en la sección anterior.

El esquema general de esta sección se muestra en la Figura 3.23. El procedimiento que se debe seguir consta de los siguientes pasos:

- Realizar una inspección visual de los estadísticos y de las funciones de densidad de probabilidades, autocorrelación simple y autocorrelación parcial. Las curvas entre las series reales y sintéticas debieran ser similares.
- Calcular los indicadores descritos en este capítulo. Cuanto menor sea el RMSE será mejor la representación estadística de la serie sintética.
- Si las series experimentaron variaciones estacionales tales como: mensuales, diarias, horarias, etc. los estadísticos deberán ser calculados para la muestra total y para sub muestras divididas de acuerdo a las variaciones estacionales. Este paso es de vital importancia en la metodología propuesta dado que a través de estos indicadores se toma la decisión en la sección siguiente.

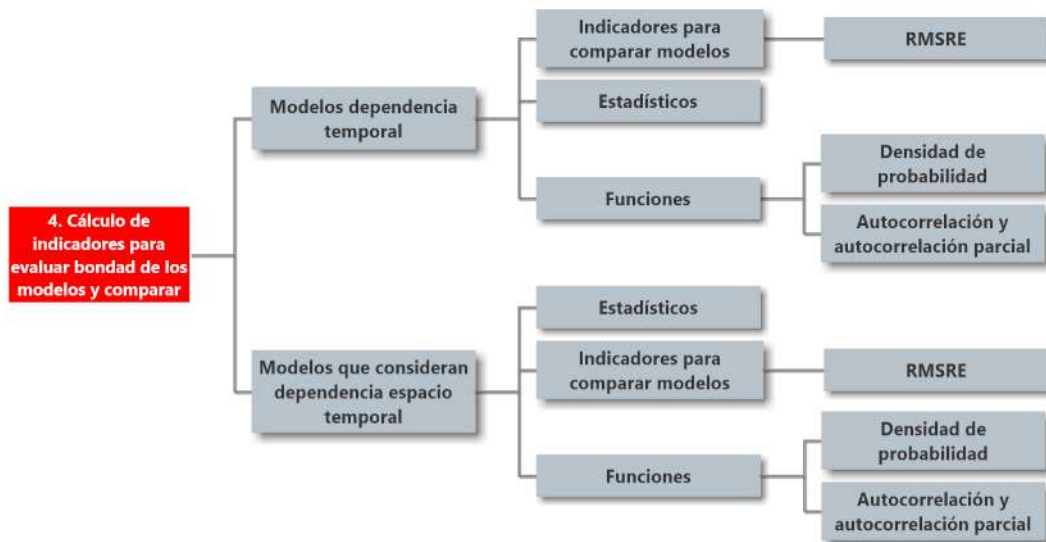


Figura 3.23. Subetapa 4: cálculo de indicadores para evaluar bondad de los modelos-esquema general

Toma de decisión del modelo

El objetivo de esta subetapa es seleccionar el modelo adecuado en función de los indicadores que son calculados en la sección anterior más la información obtenida a través de las preguntas realizadas respecto al problema de SEP. El esquema general se muestra en la Figura 3.24.

Los datos de entrada para esta sección serán los indicadores calculados para cada modelo en la sección anterior y las respuestas a las preguntas que se muestran en la Figura 3.7. El criterio de selección es en función del tipo de simulación y horizonte del problema de SEP. A continuación se describe el proceso de selección:

1. Para los problemas de muy corto plazo y corto plazo la simulación es del tipo cronológica. Por ello, es de vital importancia que los modelos para generar series sintéticas sean capaces de caracterizar

la dependencia temporal y espaciotemporal adecuadamente. Así la función de autocorrelación, función de autocorrelación parcial y matriz de autocorrelación cruzada de la serie sintética deberán ser lo más similar posible, ello se verá reflejado en el modelo que tenga el RMSE menor.

2. En los problemas de mediano, largo y muy largo plazo pueden existir variaciones dependiendo del tipo de simulación que se considere. Si la simulación es del tipo energética, lo que es usual en problemas de coordinación hidrotérmica y problemas de largo y muy largo plazo, las series sintéticas deben considerar efectos estacionales (en caso que existan) y la contribución energética. Los estadísticos son calculados para distintos intervalos de tiempo y tendrá mayor relevancia aquel modelo que mantenga un RMSE menor para los estadísticos en conjunto.
3. En los problemas de mediano, largo plazo y muy largo plazo en ocasiones se utiliza simulación cronológica pero considerando solo algunos días típicos. Este tipo de simulación exige que la estacionalidad, estructura de dependencia temporal y espacial así como la contribución energética sean caracterizadas adecuadamente. Además de caracterizar los estadísticos deberá representar las estructuras de correlación temporal y espacial así como los estadísticos para distintos intervalos de tiempo. El modelo seleccionado será aquel que posea el menor promedio ponderado del RMSE asociado a los estadísticos y la matriz de autocorrelación cruzada. De esta forma, se da igual importancia tanto a la correlación espacio temporal y a la contribución energética por estación de los modelos candidatos.

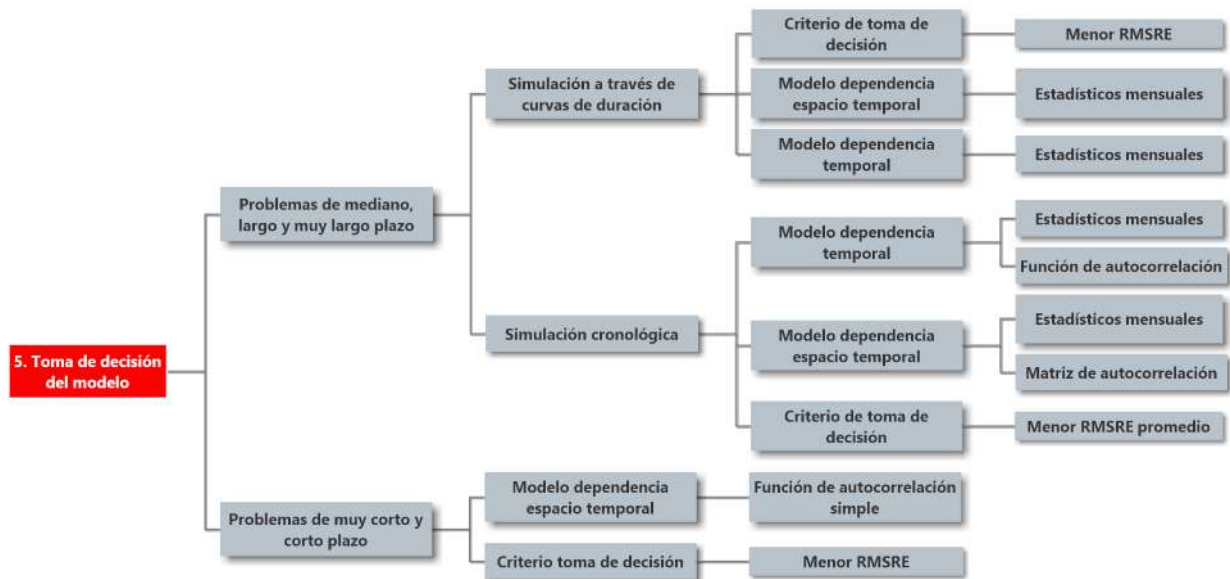


Figura 3.24. Subetapa 5: Seleccionar modelo corregir modelo

3.3 Implementación

El diagrama de flujo utilizado para la implementación de la metodología propuesta se muestra en la Figura 3.25. La implementación contempla procesos que deben ser realizados en forma manual por aquel que aplique la metodología, debido a que la completa automatización se encuentra fuera de los alcances de este trabajo de investigación.

Dos problemas de estudio se han considerado para aplicar la metodología, el primero corresponde a un problema de planificación de la operación de mediano plazo y otro de muy corto plazo.

Los datos de series de recurso que han sido utilizados en este trabajo corresponden a series de velocidad de viento obtenidos de diversos sitios de Nueva Zelanda [125] y series de potencia eólica y fotovoltaica de centrales conectadas en el sistema eléctrico interconectado central (SIC).

Los cálculos para estimar los parámetros de los modelos han sido desarrollados en dos programas. El primero corresponde a R, programa gratuito que tiene como principal enfoque el análisis estadístico, series de tiempo, entre muchas otras aplicaciones [133]. El segundo programa utilizado fue MATLAB y se utilizó para encontrar los parámetros de los modelos multivariantes. Todos los modelos VAR(p) fueron obtenidos con la librería arfit desarrollada y presentada por los autores de [134]. Además, las pruebas de hipótesis, análisis gráfico y cálculo de indicadores fueron realizados con MATLAB.

Los modelos que fueron aplicados y que serán utilizados en el siguiente capítulo son: VAR(p), CARMA(p,q), Cópula Gaussiana y PCA más ARMA(p,q).

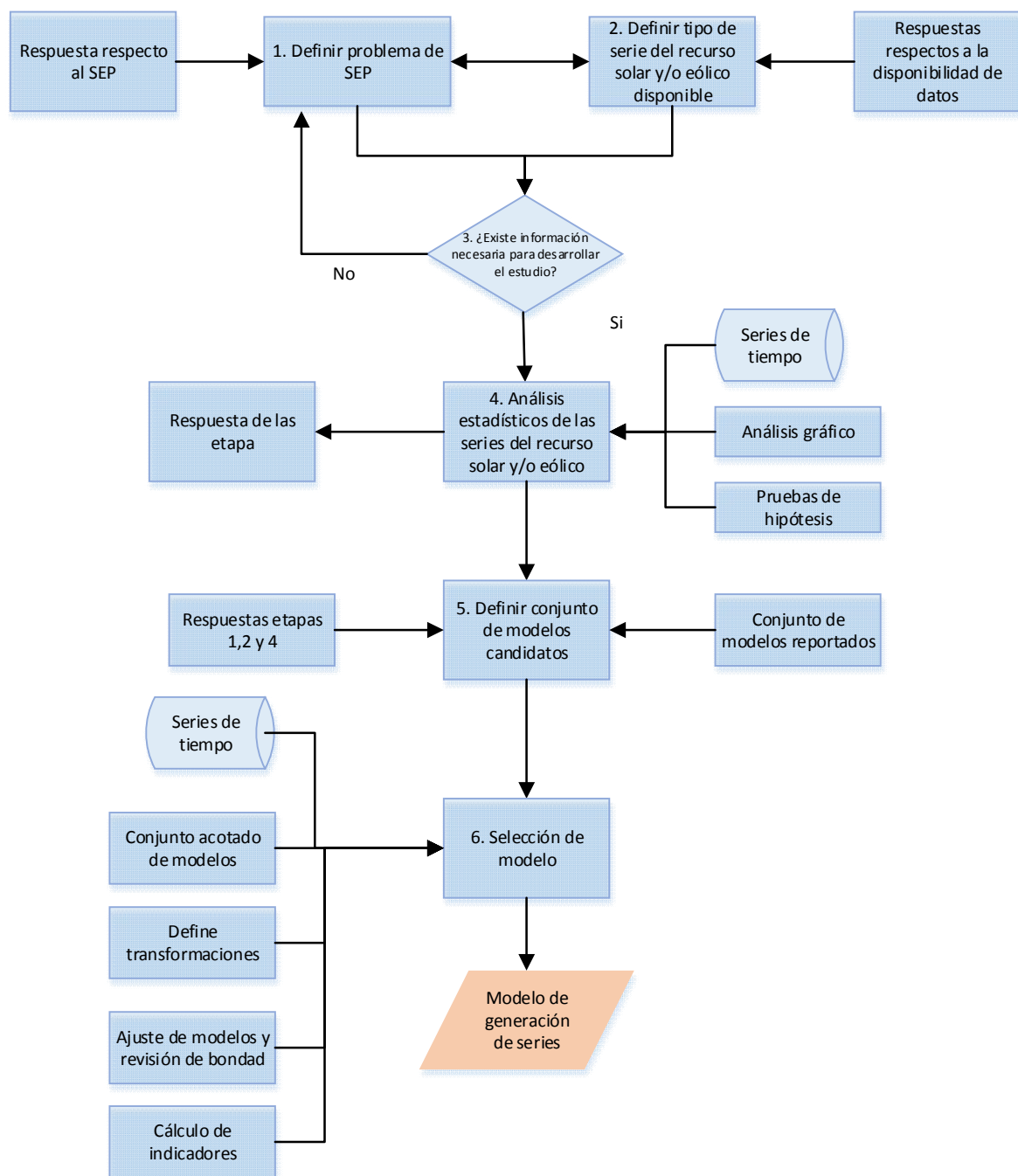


Figura 3.25. Diagrama de flujo de implementación de la metodología.

3.4 Resumen

En este capítulo se ha descrito en forma detallada la metodología que permite sistematizar la selección de modelos para generar series sintéticas. La metodología está compuesta por seis etapas y cada etapa contiene subprocesos que se deben seguir secuencialmente.

Los modelos VAR(p) a través de diversas transformaciones han demostrado ser de gran utilidad para generar series sintéticas. Debido a que existen diversas propuestas en la literatura para definir el orden adecuado de estos modelos y distintas técnicas de transformación para obtener un proceso Gaussiano y estacionario, se ha planteado un proceso que contiene los elementos relevantes para la estimación de parámetros, verificación de función de densidad de los errores y muestreo considerando correlaciones espacio temporales siendo las funciones de densidad no necesariamente normales.

El indicador que se ha definido para seleccionar modelos corresponde al RMSE y se calcula en base a los estadísticos definidos, tales como: promedio, desviación estándar, función de autocorrelación, función de autocorrelación parcial entre otros. La selección utilizando estos indicadores es basada en los objetivos que puedan tener el problema de SEP, el horizonte de evaluación y el tipo de simulación.

4 Casos de estudio y análisis de resultados

4.1 Introducción

En este capítulo se presentan dos casos de estudio. En el primer caso se aplica la metodología propuesta para un problema de planificación de la operación de mediano plazo por ser uno de los problemas en los que más se han utilizado las series sintéticas. En el segundo caso se aplica la metodología a un problema de planificación de la operación de muy corto plazo, en donde se modela series de potencia fotovoltaica y eólica buscando utilizar un modelo para generar escenarios que sirvan como datos de entrada al problema de predespacho estocástico.

En ambos casos se aplica la metodología de selección paso a paso. En el primer caso se realiza una comparación entre el modelo seleccionado con la metodología propuesta en conjunto con diversas metodologías propuestas en la literatura. Se ven los errores que pueden ocurrir al aplicar directamente una metodología para generar series sintéticas sin analizar estadísticamente las series disponibles ni el problema de SEP que se desea resolver. Por otro lado, el segundo caso muestra de que forma un modelo lineal, como lo es el modelo VAR puede ser utilizado y dar muy buenos resultados para caracterizar series con propiedades no lineales, logrando esto a través de transformaciones que han sido propuesta en la literatura especializada.

4.2 Primer caso de estudio: Modelo de planificación de la operación de mediano plazo

En esta sección se analizará la aplicación de la metodología para un problema ficticio de planificación de la operación de mediano plazo. En general se tienen los siguientes datos del problema:

- Problema de planificación de la operación de mediano plazo.
- El horizonte de evaluación corresponde a 2 años.
- El sistema contiene dos centrales eólicas y se dispone de un año de mediciones de velocidad de viento.
- Durante el tiempo de simulación no se prevé la instalación de nuevas centrales de ERV.
- El tiempo de muestreo es de una hora.
- El modelo del sistema eléctrico es simplificado, se consideran las principales líneas de transmisión y se considera una simulación cronológica.

Se aplicará la metodología propuesta etapa por etapa. Además, se realizará el análisis estadístico y la comparación de los modelos candidatos. Además, se mostrarán los resultados de aplicar indistintamente cualquier metodología para generar series sintéticas sin considerar los aspectos que han sido mencionados en la propuesta metodológica del capítulo 3.

Etapas 1: Definir el problema de SEP

Dado los datos entregados es posible contestar las siguientes preguntas de esta etapa

1. ¿Qué tipo de problema de sistema eléctrico se desea resolver?

Planificación de la operación mediano plazo

2. ¿Qué tipo de simulación se utilizará?

Simulación cronológica

3. ¿Qué tipo de modelo del sistema eléctrico es considerado?

Sistema multinodal

4. ¿Cuál es el orden del horizonte de evaluación del problema de SEP?

Años

5. ¿Cuál es el tiempo mínimo de simulación?

Hora (considera tiempos t , tal que $1 \text{ min} < t \leq 60 \text{ minutos}$)

6. ¿Cuáles son los requerimientos de los modelos para centrales renovables?

Modelar sólo plantas existentes

Las respuestas se almacenan y vuelven a ser utilizadas en las etapas 5 y 6.

Etapla 2: Definir tipo de serie del recurso solar y/o eólico disponible

Se cuentan con mediciones de velocidad de viento en los lugares de interés. Las mediciones de velocidad consideradas para realizar el análisis estadístico y ajustar los modelos corresponden a sitios ubicados en Nueva Zelanda [125]. En la Figura 4.1 se muestra la ubicación y los nombres de cada sitio. En el caso se utilizaron los sitios STH1 y CKS1 cuya elección fue realizada en forma arbitraria. Los registros pueden ser accedidos a través del sitio personal del investigador Geoffrey Pritchard perteneciente a la Universidad de Auckland [135].

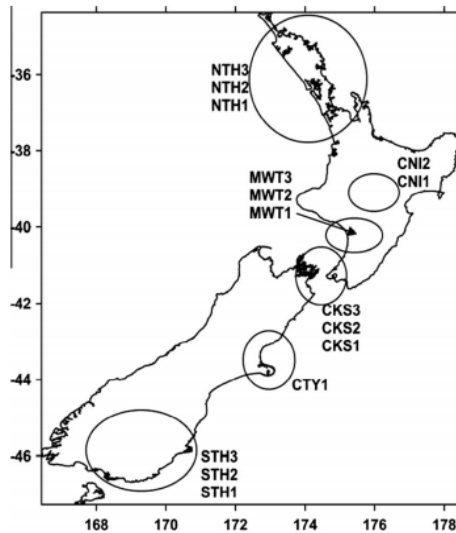


Figura 4.1. Ubicación de mediciones en Nueva Zelanda [125]

Considerando la descripción del problema ya es posible contestar las preguntas de esta etapa

1. ¿Cómo se generaron las series con registros de los recursos renovables?

Mediciones del recurso

2. ¿Cuál es el recurso del que se tiene información?

Velocidad de viento a la altura de las turbinas eólicas

3. ¿Cuál es la resolución temporal de los registros?

Hora (considera tiempos t , tal que $1 \text{ min} < t \leq 60 \text{ minutos}$)

Las respuestas son almacenadas para ser utilizadas en las etapas 5 y 6.

Etapla 3: ¿Existe información necesaria para desarrollar el estudio?

El problema considera simulación cronológica, cuyo tiempo de simulación es de una hora. Por otro lado, las mediciones también tienen un tiempo de muestreo de una hora, por lo tanto, existe coherencia entre el estudio y las mediciones, y con ello es posible pasar a la siguiente etapa.

Etapla 4: Análisis estadístico de las series del recurso solar y/o eólico

Se aplica el proceso de la Figura 3.14 a las series de tiempo de viento que se disponen. La información general de ellas se muestra a continuación:

- Tiempo de muestreo: 1 hora
- Longitud de la muestra: 8737 mediciones
- Fecha: desde el 01-01-2008 hasta el 31-10-2008
- Los sitios analizados son SH1 y CKS1
- Se deben responder las siguientes preguntas en la etapa 4
 - ¿Qué tipo de función de densidad de probabilidades tienen los recursos?
 - ¿Tiene alguna de las series tendencia?
 - ¿Tiene alguna de las series variaciones estacionales?
 - ¿Existe dependencia temporal?
 - ¿Existe dependencia espacial?
 - ¿Son todas las series estacionarias?

A continuación se contestan estas preguntas aplicando los métodos gráficos y estadísticos propuestos en el capítulo 3.

¿Qué tipo de función de densidad de probabilidades tienen los recursos?

Realizando una inspección visual de los histogramas que se muestran en la Figura 4.2 es posible descartar que se distribuyan forma normal. Sólo a modo de ejemplo se aplican las pruebas de hipótesis χ^2 y Anderson-Darling. Para ambas pruebas los p – valores ≈ 0 , lo que implica que se rechaza la hipótesis nula de normalidad.

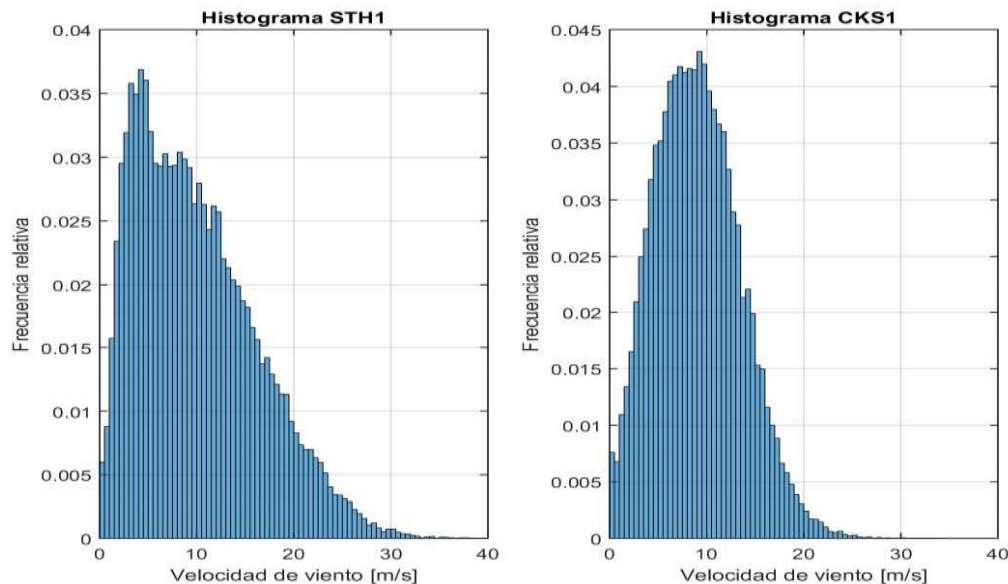


Figura 4.2. Histogramas de las series de velocidad de viento en los sitios STH1 y CKS1

¿Tiene alguna de las series tendencia?

Según la gráfica que se muestra en la Figura 4.3 no se observa una tendencia en las series de tiempo, lo que es usual en las series de velocidad de viento. Además, es posible aplicar la prueba de hipótesis Dicker-Fuller aumentada, la cual busca verificar si existe alguna raíz unitaria en la serie temporal, versus la hipótesis alternativa de ajuste de un proceso ARMA. La prueba de hipótesis rechaza la hipótesis nula, por lo tanto no se observa tendencia marcada en la serie.

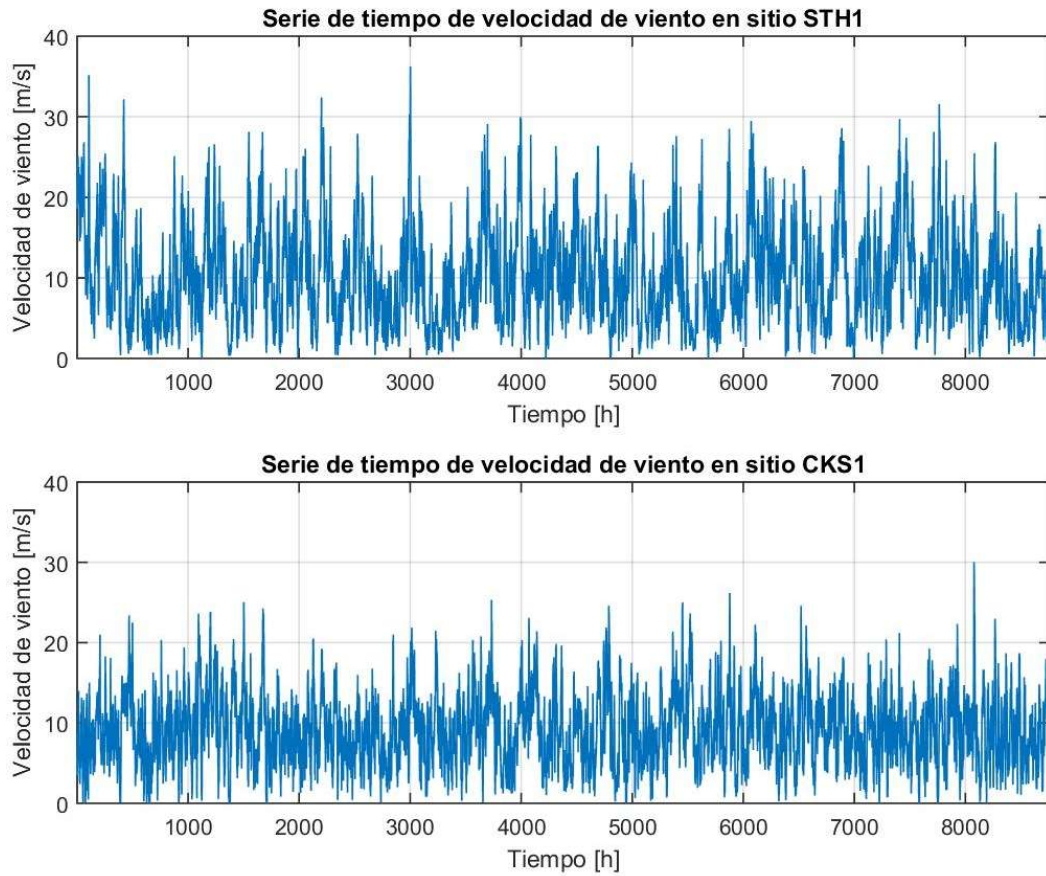


Figura 4.3. Series de tiempo de tiempo de velocidades de viento para los sitios STH1 y CKS1

¿Tiene alguna de las series variaciones estacionales?

Para observar si existen variaciones estacionales se grafican en la Figura 4.4 los diagramas de cajas para las variaciones horarias y mensuales. En ambos sitios es posible observar que existe una variación estacional horaria, observándose un patrón notorio en la media durante el día. Por otro lado, para el caso mensual no se observa una variación muy marcada, por lo que a través del diagrama de caja es difícil establecer la existencia de estacionalidades mensuales. Para verificar la existencia de estacionales horarias y mensuales, se aplica la prueba de hipótesis Kruskal Wallis. En ambos casos y sitios los $p - valores \approx 0$, lo que implica que existen variaciones estacionales horarias y mensuales.

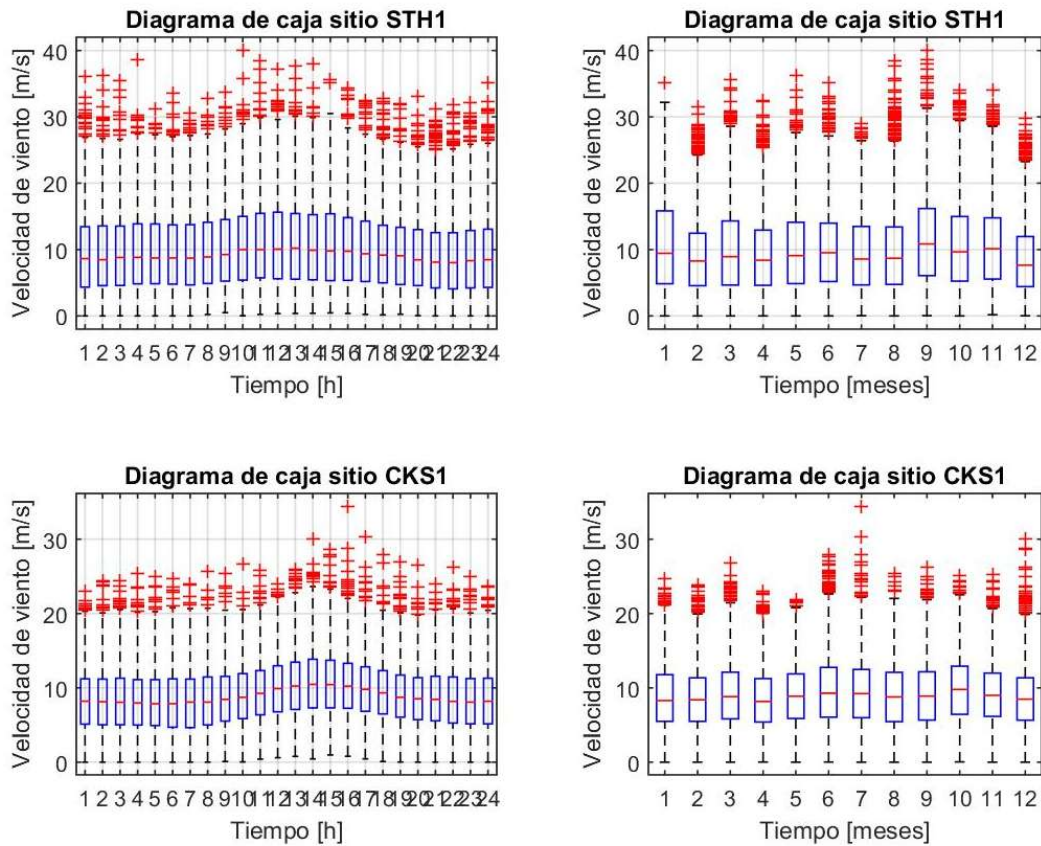


Figura 4.4. Diagrama de cajas para variaciones horarias y mensuales para los sitios STH1 y CKS1

¿Existe dependencia temporal?

La dependencia temporal es observada a través de las funciones de autocorrelación simple (ACF) y parcial (PACF). En la Figura 4.5 se muestran dichas funciones. Las líneas azules de estas gráficas representan límites de confianza. Dado que existen coeficientes de autocorrelación mayores a esos límites se puede concluir que las series de tiempo se encuentran correlacionadas temporalmente, lo cual se encuentra dentro de lo esperado para estos recursos. Más análisis no es necesario, con estas funciones es suficiente.

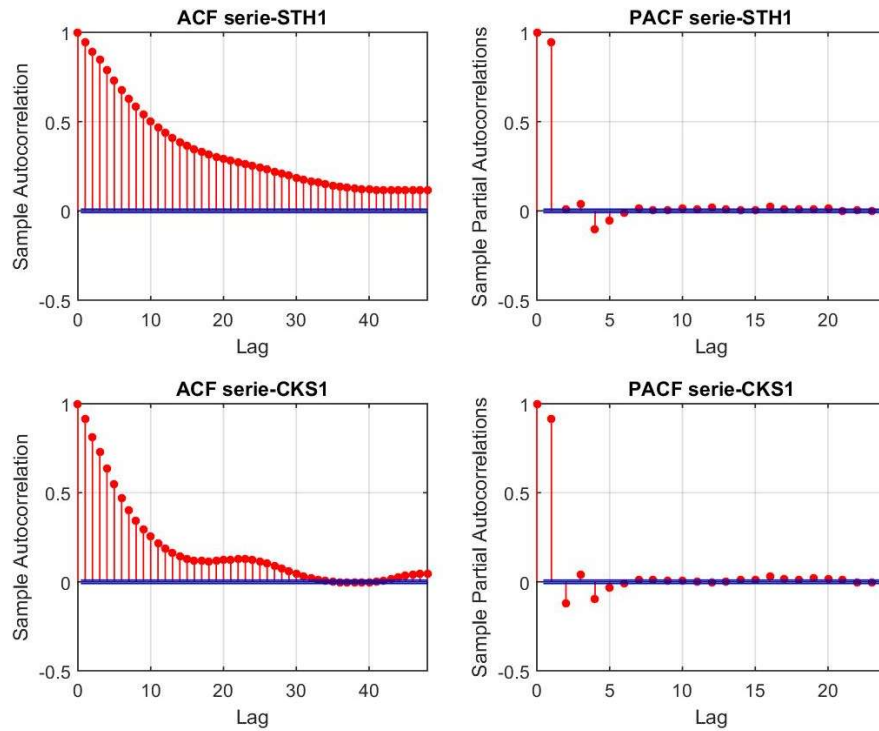


Figura 4.5. Funciones de autocorrelación simple y parcial correspondientes a los sitios STH1 y CKS1

¿Existe dependencia espacial?

El análisis de la dependencia espacial se realiza a través de la prueba de hipótesis Ljung Box en su versión multivariable, en especial cuando el número de series de tiempo es mayor a dos. En la Figura 4.6 se muestra que los p-valores son menores al nivel de confianza del 5%, ello implica que se rechaza la hipótesis nula de que la matriz de autocorrelación es cero. Esta conclusión se puede corroborar a través de los gráficos de la función de autocorrelación cruzada y los diagramas de dispersión que se muestran en las Figura 4.7 y Figura 4.8 respectivamente. El coeficiente de correlación entrega información respecto al segundo momento de la función de densidad de probabilidades y el diagrama de dispersión entrega información respecto a la estructura de dependencia que existe entre los sitios bajo estudio.

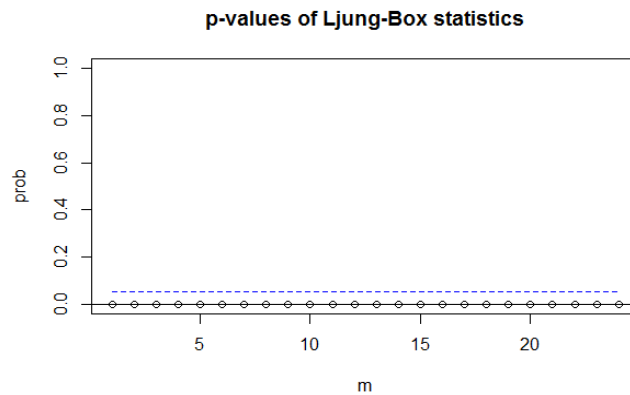


Figura 4.6. Gráfico de p-valores de la prueba de hipótesis Ljung box para modelos multivariables

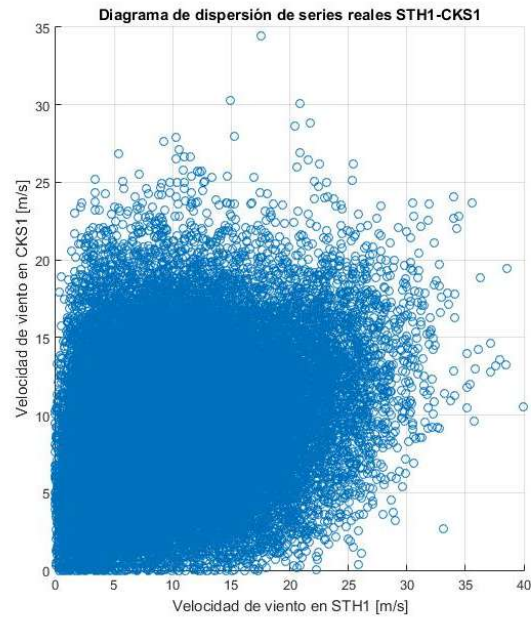


Figura 4.7. Diagrama de dispersión entre los sitios STH1 y CKS1

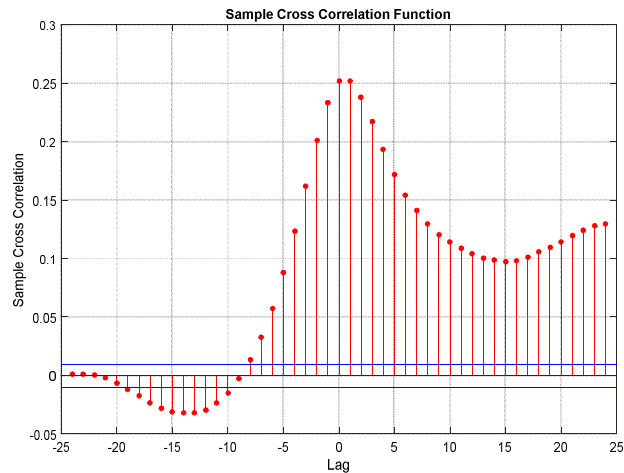


Figura 4.8. Función de autocorrelación cruzada

¿Son las series estacionarias?

Debido a la existencia de estacionalidades las series no son estacionarias. En resumen en esta etapa se tienen las siguientes respuestas al cuestionario:

1. ¿Qué tipo de función de densidad de probabilidades tienen los recursos?

No Gaussiana

2. ¿Tiene alguna de las series tendencia?

No

3. ¿Tiene alguna de las series variaciones estacionales?

Sí

4. ¿Existe dependencia temporal?

Sí

5. ¿Existe dependencia espacial?

Si

6. ¿Son todas las series estacionarias?

No

Estas respuestas son almacenadas y utilizadas en las etapas siguientes de la metodología propuesta.

Etapas 5: Definir conjunto de modelos candidatos

Siguiendo el proceso descrito en el apartado 3.2.5 y el diagrama de flujo de la Figura 3.17 es posible obtener los modelos candidatos. Para este caso se tienen dos series de viento con igual cantidad de datos sincronizados, correlacionados espacial y temporalmente, con estacionalidad horaria, mensual y finalmente no gaussiano. Para estas características se busca en la base de datos del anexo 4, entre los modelos que consideran las características de las series de tiempo que se tienen disponible. Los modelos 17, 18, 20, 23 y 26 de la base de datos podrían ser utilizados de acuerdo a las características del problema y de las series de tiempo.

Estos modelos utilizan las siguientes estructuras matemáticas:

- VAR(p): Corresponde a las metodologías 17, 18 y 20
- CARMA(p,q) considerando coeficiente de correlación cruzada: corresponde al modelo 23
- PCA+ ARMA(p,q) : corresponde al modelo 26

Los modelos CARMA más simples no se consideran porque los datos se encuentran sincronizados y se comentó con anterioridad que la simplificación de este modelo ocasionaba que la correlación espacial no se pueda considerar adecuadamente. La teoría de cópula no puede ser ocupada adecuadamente porque sólo se dispone de un año de mediciones. Debido a que el problema de SEP es de mediano plazo y las series contienen estacionalidades, un año de datos no es suficiente para caracterizar las variaciones estacionales.

Dado los alcances de esta tesis, se aplican los modelos VAR con sus variantes y PCA+ARMA. Estos modelos se encuentran detallados en el anexo 2. También se encuentra en dicho anexo el ajuste de los modelos y el análisis de la bondad de ellos.

Los modelos son almacenados y se pasa a la siguiente etapa.

Etapas 6: Selección de modelo

Los datos de entrada a esta etapa son las series de tiempo de mediciones y los modelos candidatos de la etapa anterior. Esta etapa consta de los siguientes procesos:

1. Determinación de procesamiento previo
2. Determinación de parámetros de los modelos
3. Generación de series sintéticas
4. Cálculo de indicadores para evaluar la bondad de los modelos
5. Toma de decisión del modelo

Dado que un modelo candidato corresponde a un proceso VAR, se aplica el diagrama de flujo de la Figura 3.21 para ajustar sus parámetros y determinar el procesamiento y transformaciones necesarias.

Como se ha mencionado en el estado del arte, los procesamientos que pueden ser realizados para quitar la estacionalidad son aquellos mostrados en las siguientes ecuaciones:

$$Y_t = X_t - X_{at} - X_{dt} \quad (3.13)$$

$$Y_t = \frac{X_t - u_{tk}}{\sigma_{tk}}, \quad \forall k \in \{S\} \quad (3.14)$$

$$Y'_t = F_{invN}^{-1}(F_k(Y_{tk})), \quad \forall k \in \{S\}$$

$$\begin{aligned} Y_t &= X_t - u_{tk}, \quad \forall k \in \{S\} \\ Y'_t &= F_{invN}^{-1}(F(Y_t)) \end{aligned} \quad (3.15)$$

De acuerdo al diagrama de flujo, se debe aplicar cada procedimiento para transformar la serie real a una en la cual se pueda aplicar la estructura VAR (p). Una vez aplicada las diferentes transformaciones se aplican las mismas pruebas de hipótesis de la etapa de análisis estadístico a la serie transformada. Se debe verificar que la estacionalidad se haya removido y la serie resultante sea gaussiana con media $\mu = 0$ y desviación estándar $\sigma = 1$.

Para verificar que la metodología propuesta sistematiza el proceso de selección, se utilizaron otras metodologías y modelos para generar series sintéticas las cuales se muestran en la

Tabla 4.2 y la Tabla 4.3. Las filas que se encuentran en color celeste corresponden a los modelos que son candidatos y el resto son distintas combinaciones de las ecuaciones (3.13)-(3.15) las cuales se describen en la columna “Transformaciones” y han sido desarrolladas en diferentes artículos de investigación según ha sido presentado en la

Tabla 4.2 y Tabla 4.3. La estimación del orden de cada modelo es realizado utilizando el criterio AIC (Para mayor detalle ver anexo 1, el cual contiene información respecto a la definición del criterio AIC y ver el anexo 2 para ver un ejemplo de la utilización del criterio AIC en la estimación del orden del modelo VAR)

De los modelos candidatos se tiene que el modelo 1 no pasa la prueba de normalidad, por lo tanto el procesamiento propuesto en dicha metodología no es apto para el caso que se está analizando. Los modelos 8, 9 y 14 pasan las pruebas de hipótesis de normalidad, pero el modelo 8 no pasa la prueba de hipótesis de Kruskal Wallis para las variaciones mensuales, por ende no sería estacionaria la serie transformada y no sería apropiado utilizar este modelo.

La estimación del orden del modelo VAR (p) y PCA+ARMA (p, q) se realiza utilizando AIC en su versión multivariable para el modelo VAR (p) y en univariable para el modelo PCA+ARMA (p, q) (En el anexo2 se detalla el algoritmo y resultados aplicándolos a las series que se estudian en este caso). El orden se muestra en la

Tabla 4.2 y la Tabla 4.3. Luego se calculan los parámetros y se aplican las pruebas de hipótesis de normalidad y de autocorrelación a los residuos. Estas pruebas muestran que se cumplen con los supuestos. El resumen de los resultados se muestra en la Tabla 4.1

Tabla 4.1. Resumen de análisis a series transformadas y residuos de los modelos propuestos

Tipo modelo	Orden del modelo	Sitio	Normal	Estacionaria	Estacional	ACF residuos	CCF residuos	Normal residuos
Modelo 1	VAR(16)	STH1	No	Si	No	No	No	No
Modelo 1	VAR(16)	CKS1	No	Si	No	No	No	No
Modelo 8	VAR(16)	STH1	Si	No	Si	No	No	Si
Modelo 8	VAR(16)	CKS1	Si	No	Si	No	No	Si
Modelo 9	VAR(16)	STH1	Si	Si	No	No	No	Si
Modelo 9	VAR(16)	CKS1	Si	Si	No	No	No	Si
Modelo 14	PCA+ARMA	STH1	Si	Si	No	No	No	Si
Modelo 14	PCA+ARMA	CKS1	Si	Si	No	No	No	Si

Tabla 4.2. Metodologías implementadas (Parte 1)

Nº	Tipo de dependencia	Transformaciones y/o procesamientos previos	Estructura del modelo	Orden del modelo
1	Espacio temporal	Resta media mensual, diaria y horaria	VAR(p)	VAR(16)
2	Espacio temporal	Resta media mensual, diaria y horaria + muestreo usando cópula Gaussiana	VAR(p)	VAR(16)
3	Espacio temporal	Transformada integral para 24 horas	VAR(p)	VAR(16)
4	Espacio temporal	Transformada integral 24 horas + muestreo usando Cópula	VAR(p)	VAR(16)
5	Espacio temporal	Resta media horaria mensual y luego se aplica transformada integral para 24 horas	VAR(p)	VAR(16)
6	Espacio temporal	Resta media horaria mensual y luego se aplica transformada integral para 24 horas+ muestreo usando Cópula	VAR(p)	VAR(16)
7	Espacio temporal	Resta media horaria mensual (288 promedios). Después aplica transformada integral y por último ajusta modelo	VAR(p)	VAR(16)
8	Espacio temporal	Resta media horaria mensual (288 promedios). Después aplica transformada integral y por último ajusta modelo + muestreo usando teoría de Copula	VAR(p)	VAR(16)
9	Espacio temporal	Resta media y div desviación estándar más transformada integral para 24 horas	VAR(p)	VAR(16)
10	Espacio temporal	Resta media y div desviación estándar más transformada integral para 24 horas+ muestreo usando teoría de Copula	VAR(p)	VAR(16)

Tabla 4.3. Metodologías implementadas (Parte 2)

Nº	Tipo de dependencia	Transformaciones y/o procesamientos previos	Estructura del modelo	Orden del modelo
11	Espacio temporal	Resta media horaria mensual y división por desviación estándar horaria mensual. Se aplica transformada integral a la serie completa	VAR(p)	VAR(16)
12	Espacio temporal	Resta media horaria mensual y división por desviación estándar horaria mensual. Se aplica transformada integral a la serie completa + cópula	VAR(p)	VAR(16)
13	Espacio temporal	Resta media horaria mensual y división por desviación estándar horaria mensual. Se obtiene un modelo ARMA(p,q) distinto para cada sitio. Se simula y luego se multiplica por la matriz obtenida a través de la factorización de Cholesky	CARMA(p,q)	CARMA(3,3) Para ambos sitios
14	Espacio temporal	Resta media horaria mensual y división por desviación estándar horaria mensual. Se aplica PCA para finalmente encontrar modelo ARMA.	PCA+ARMA(p,q)	PCA+ARMA(20,3) Para STH1 y PCA+ARMA(20,4) CKS1
15	Espacio temporal	Teoría de Cópula. Se utiliza una cópula gaussiana	Cópula Gaussiana	No aplica
16	Espacio temporal	Resta media horaria mensual y división por desviación estándar horaria mensual. Se obtiene un modelo ARMA(p,q) distinto para cada sitio y luego se muestrea usando factorización de Cholesky	CARMA(p,q)	CARMA(3,3) Para ambos sitios
17	Espacio temporal	Resta media horaria mensual y división por desviación estándar horaria mensual. Se obtiene un modelo ARC(p,q) distinto para cada sitio y luego se muestrea usando factorización de Cholesky	CARMA(p,q)	CARMA(16,0) Para STH1 y CARMA(5,0)

Generación de series sintéticas-Cálculo de indicadores para evaluar bondad de los modelos

Se generan series sintéticas de igual longitud que las series originales y se procede a calcular los estadísticos e indicadores para verificar la bondad de los modelos y tomar una decisión.

Cálculo de indicadores para evaluar la bondad de los modelos

Debido a que el problema es de mediano plazo y para un horizonte de 2 años, se calculan los estadísticos mensuales y el indicador RMSRE. Los resultados de los indicadores se muestran en la siguiente subetapa.

Toma de decisión del modelo

En los problemas de mediano, largo y muy largo plazo en ocasiones se utiliza simulación cronológica pero considerando solo algunos días típicos. Este tipo simulación exige que la estacionalidad, estructura de dependencia temporal y espacial, así como la contribución energética sea caracterizada adecuadamente. El modelo, además de caracterizar los estadísticos, deberá representar las estructuras de correlación temporal y espacial, así como los estadísticos para distintos intervalos de tiempo. El modelo seleccionado será aquel que posea el menor promedio ponderado del RMSE asociado a los estadísticos y a la matriz de autocorrelación cruzada. De esta forma, se da igual importancia tanto a la correlación espacio temporal como a la contribución energética por estación de los modelos candidatos.

El problema de SEP considera simulación cronológica, y es de mediano plazo, específicamente 2 años. Debido a que las series de tiempo experimentan variaciones estacionales mensuales, se escoge el modelo que tenga el menor \overline{RMSRE} que viene dado por la ecuación:

$$\overline{RMSRE} = \frac{1}{2} (RMSRE_{EstadxMes} + RMSRE_{CCF}) \quad (3.16)$$

En la Figura 4.9 se muestra el indicador RMSRE aplicado a la estadística mensual y a la función de autocorrelación cruzada (CCF). En la Figura 4.10 se muestra la gráfica con la aplicación de la ecuación (3.16) y es posible concluir que el modelo con menor \overline{RMSRE} es el número 9.

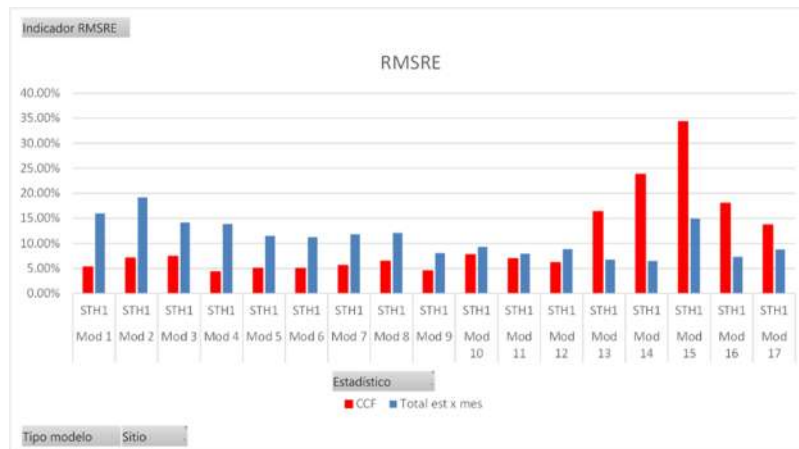


Figura 4.9. RMSRE para CCF y estadísticos mensuales

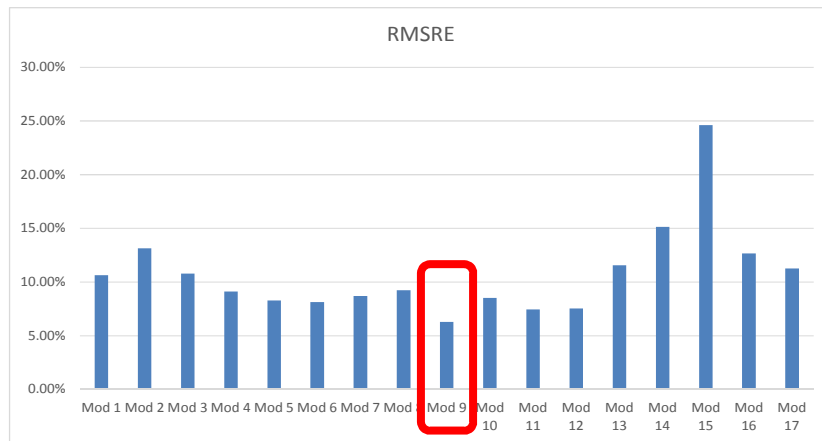


Figura 4.10. RMSRE promedio ponderado

Es importante analizar algunos aspectos que justifican el indicador y la propuesta metodológica de este trabajo. Es muy complejo tomar la decisión del modelo que se debe utilizar en función de los estadísticos aplicados a la totalidad de la serie. Para todas las metodologías implementadas, se muestra en la Figura 4.11 que el RMSRE aplicado sobre los estadísticos considerando toda la serie de tiempo, en ningún caso es observa un RMSRE mayor al 4%. Si se utilizara este indicador para seleccionar el modelo con el menor RMSRE se escogería la metodología 15 según se muestra en la Figura 4.11. Sin embargo, este modelo es el que peor caracteriza la dependencia espacio-temporal llegando a tener un RMSRE del 35% (ver Figura 4.9).

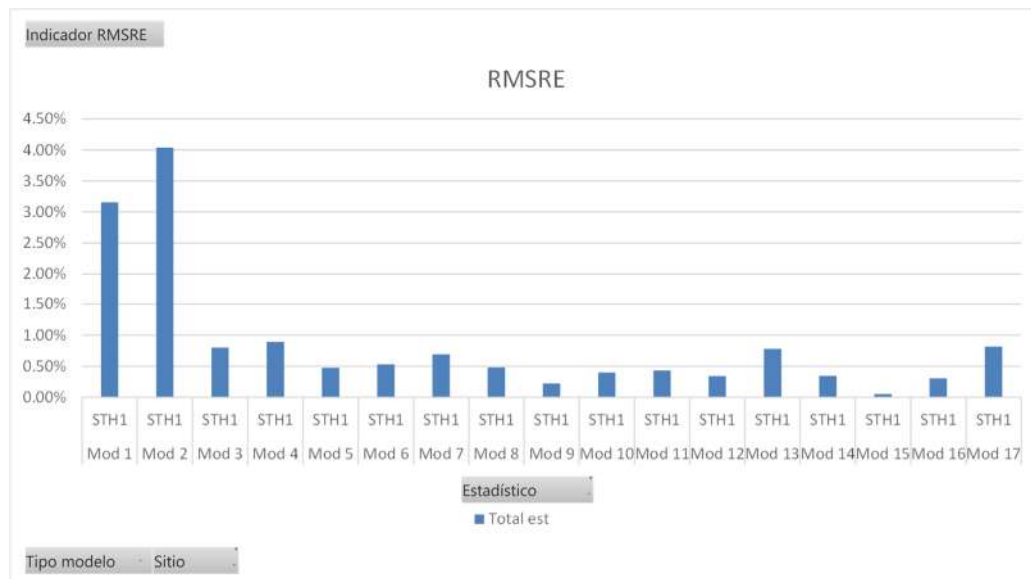


Figura 4.11. RMSRE aplicado sobre estadísticos total al sitio STH1

4.2.1 Resultados gráficos del modelo seleccionado

En esta sección se muestran los resultados que entrega el modelo seleccionado con la metodología propuesta. Se realizan comparaciones gráficas, debido a que a través de ellas es posible apreciar fácilmente la bondad del modelo.

En la Figura 4.12 y Figura 4.13 se muestra una comparación de los histogramas. De acuerdo a dichas figuras es posible observar que el modelo es capaz de caracterizar adecuadamente la función de densidad de probabilidades, aspecto importante para el ejercicio de planificación de mediano plazo. Una buena

representación de la función de densidad implica que las contribuciones energéticas consideradas en el problema de SEP serán adecuados. Además, dado que el problema considera realizar una simulación cronológica, las contribuciones energéticas deben estar en armonía con la estacionalidad que manifiesta la serie real. Desde la Figura 4.20 a la Figura 4.23 se comparan los diagramas de caja para las variaciones horarias y mensuales. La ventaja de estos diagramas radica en el hecho que es posible observar la dispersión y distribución de los datos. En dichas figuras claramente es apreciable que el modelo para generar series sintéticas se distribuye en forma similar a la serie de tiempo real y con ello el modelo asegura que las contribuciones energéticas estén en acorde con la estacionalidad del recurso.

De acuerdo al análisis estadístico se observa que las series se encontraban correlacionadas espacio-temporalmente. Para verificar que el modelo haya generado series sintéticas que sean capaces de caracterizar estas cualidades de las series reales, se graficaron las funciones ACF, PACF y CCF en la Figura 4.14, Figura 4.15 y Figura 4.17 respectivamente. Se observa que existe similitud entre las funciones de la serie real y la sintética y sus diferencias son pequeñas. Además en la Figura 4.16, Figura 4.18 y Figura 4.19 se muestran diagramas de dispersión. Recordando que los coeficientes de correlación corresponden a los segundos momentos de la función de distribución y sólo a eso, los diagramas de dispersión entregan información de la estructura de dependencia, las cuales también quedan bien caracterizadas con el modelo propuesto.

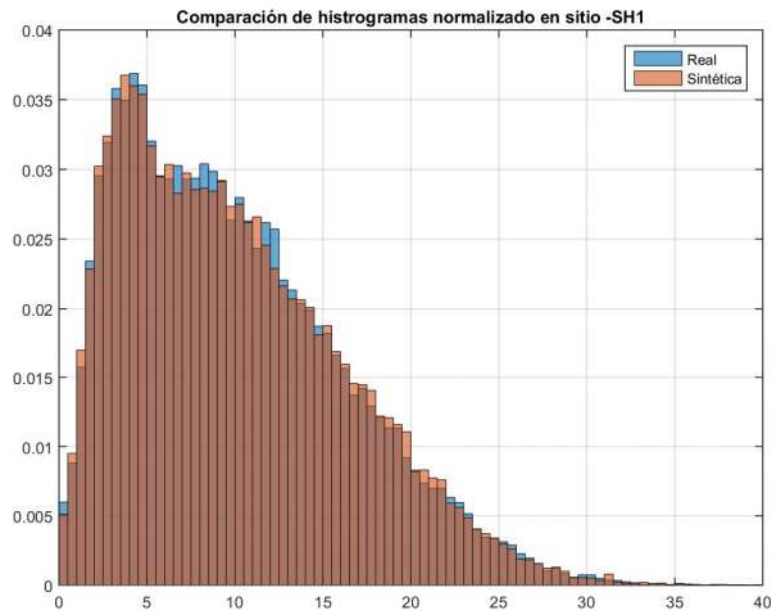


Figura 4.12. Comparación de histograma normalizado para sitio STH1

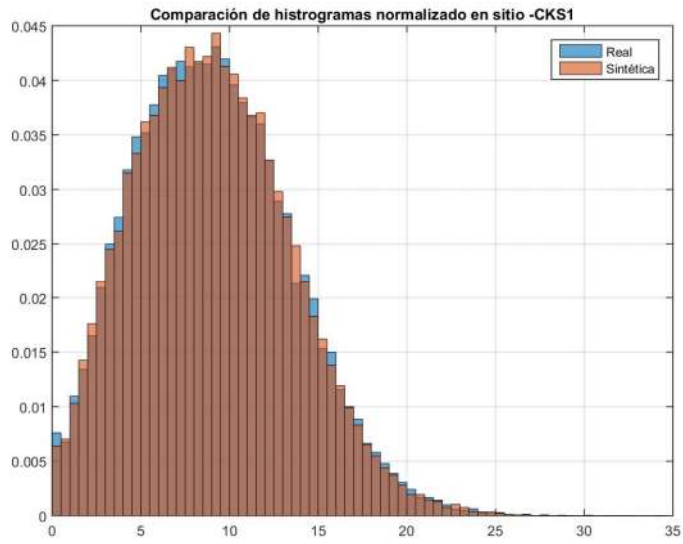


Figura 4.13. Comparación de histograma normalizado para sitio CKS1

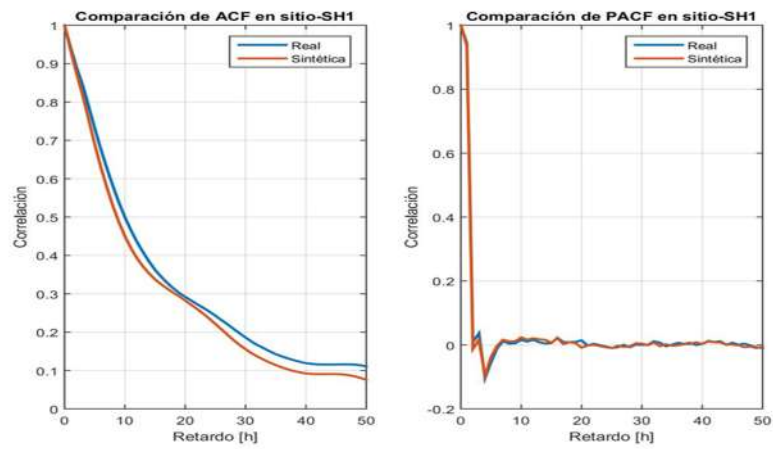


Figura 4.14. Comparación de ACF y PACF para el sitio STH1

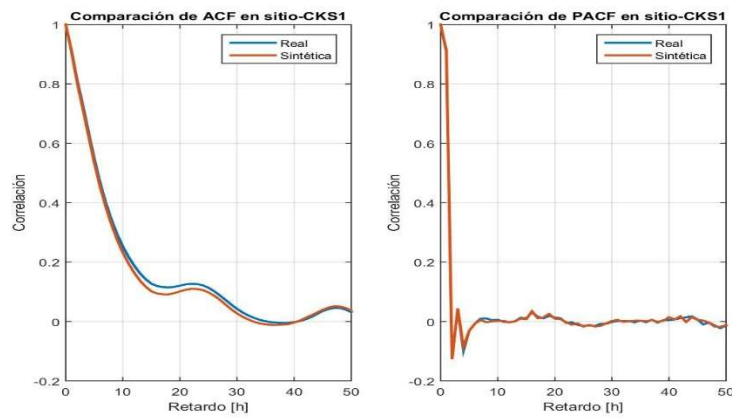


Figura 4.15. Comparación de ACF y PACF para el sitio STH1

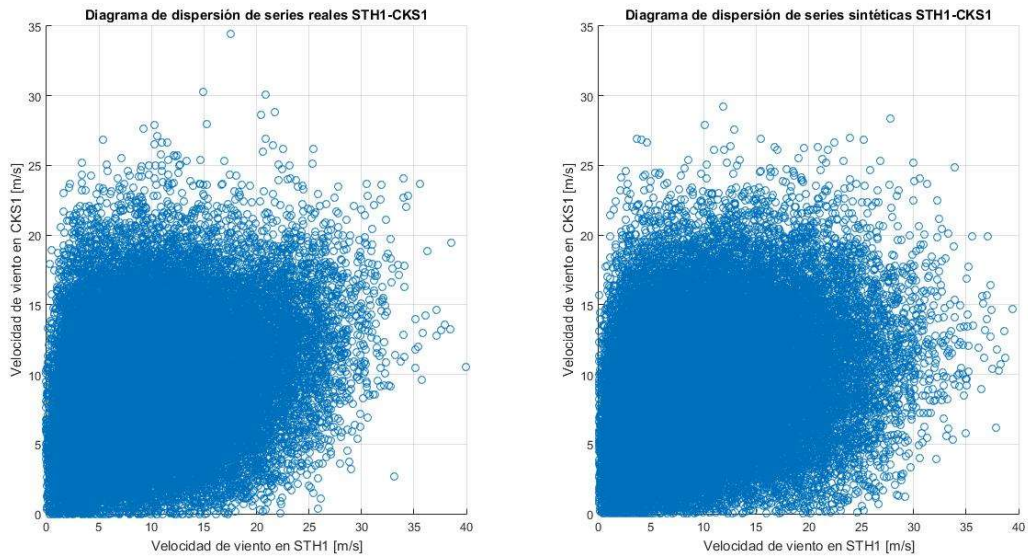


Figura 4.16. Comparación de diagramas de dispersión

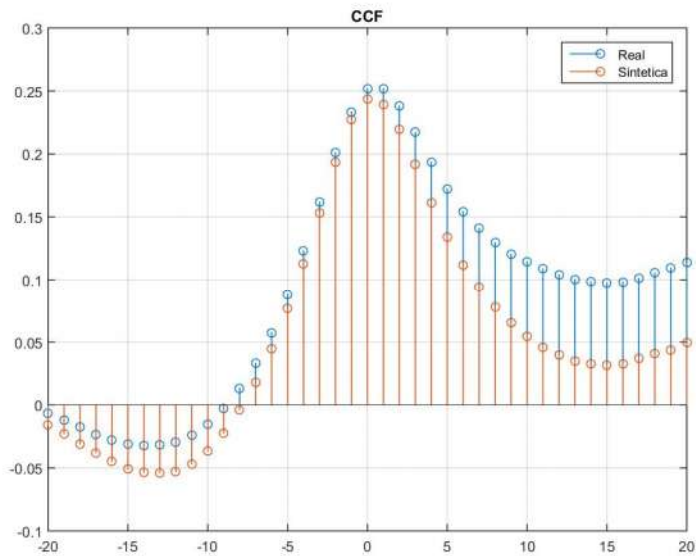


Figura 4.17. Comparación de función de autocorrelación cruzada

Las siguientes figuras muestran la correlación entre $STH1(t)$ y $STH1(t-2)$, y $CKS1(t)$ y $CKS1(t-2)$. En ambos casos se compara la serie sintética con la real y es posible observar que la serie sintética es capaz de representar adecuadamente la estructura de dependencia temporal. La estructura de dependencia es más importante que las autocorrelaciones, debido a que estas sólo representan el segundo momento de la función de densidad de probabilidades.

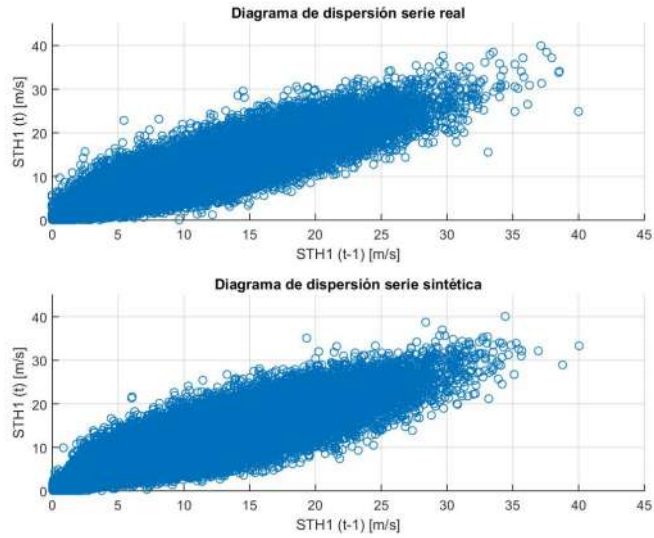


Figura 4.18. Comparación de diagramas de dispersión para el sitio STH1

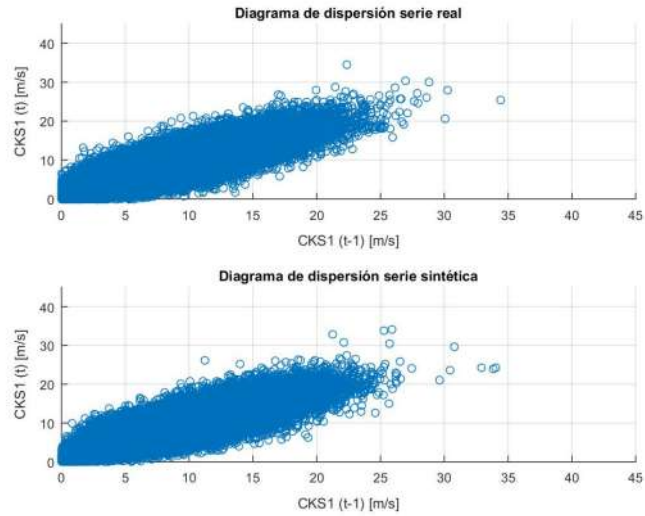


Figura 4.19. Comparación de diagramas de dispersión para el sitio CKS1

Las variaciones horarias se caracterizan adecuadamente, según se observan en los diagramas de cajas. Las variaciones de la media y diferencia intercuartil siguen el mismo comportamiento que en las series reales.

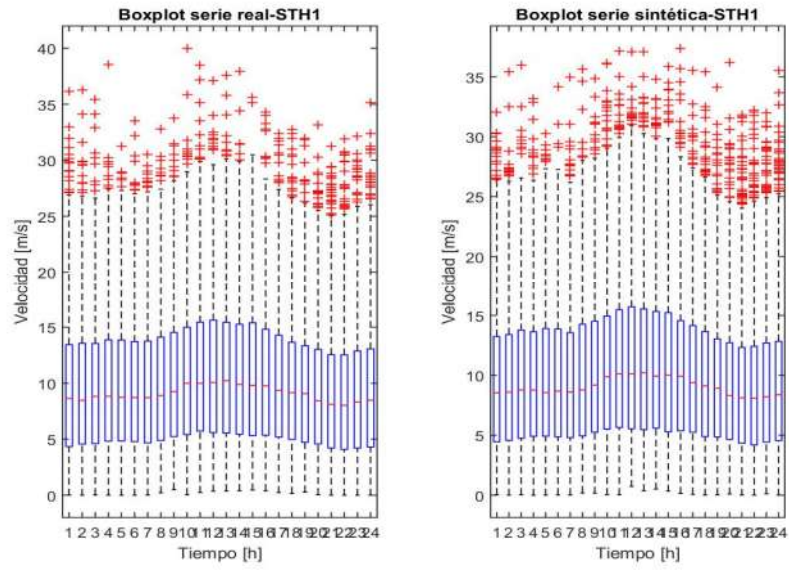


Figura 4.20. Comparación de diagrama de cajas horarios asociados al sitio STH1

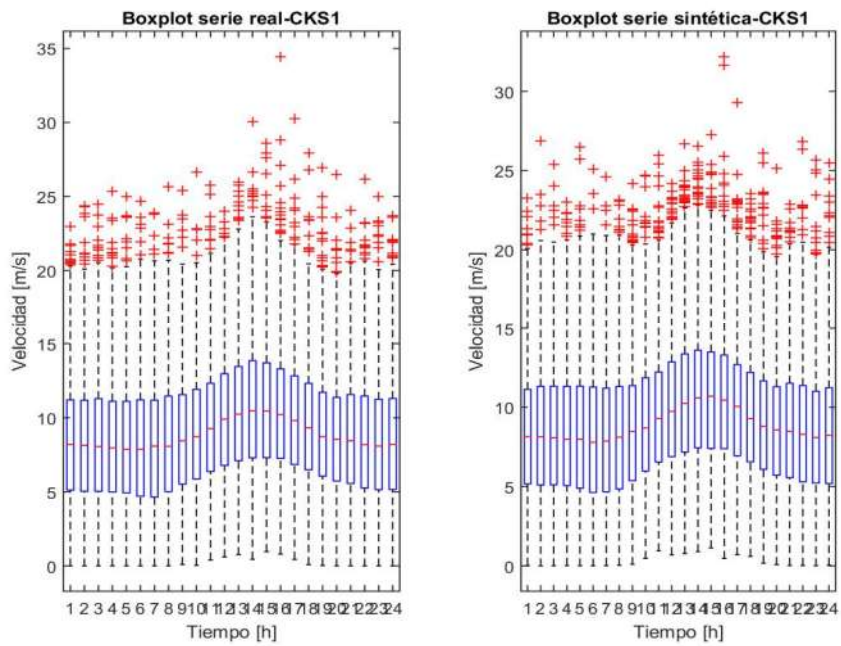


Figura 4.21. Comparación de diagrama de cajas horarios asociados al sitio CKS1

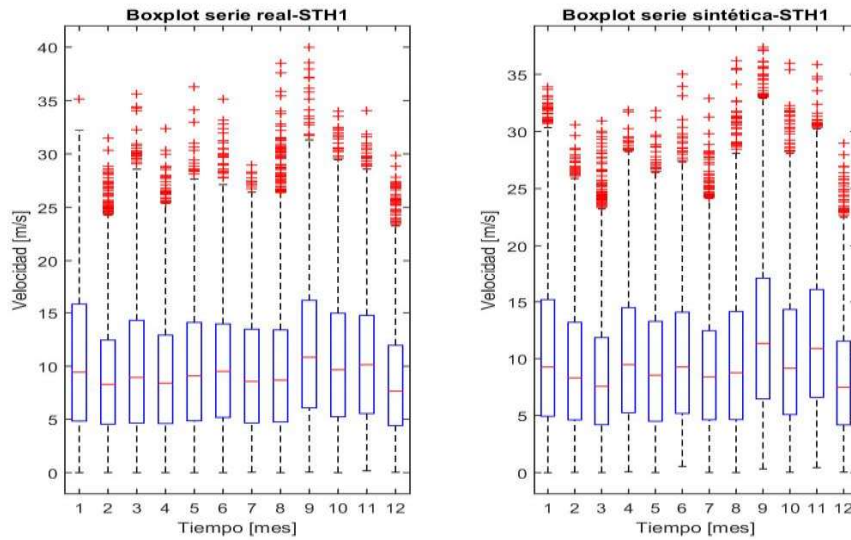


Figura 4.22. Comparación de diagrama de cajas mensuales asociados al sitio STH1

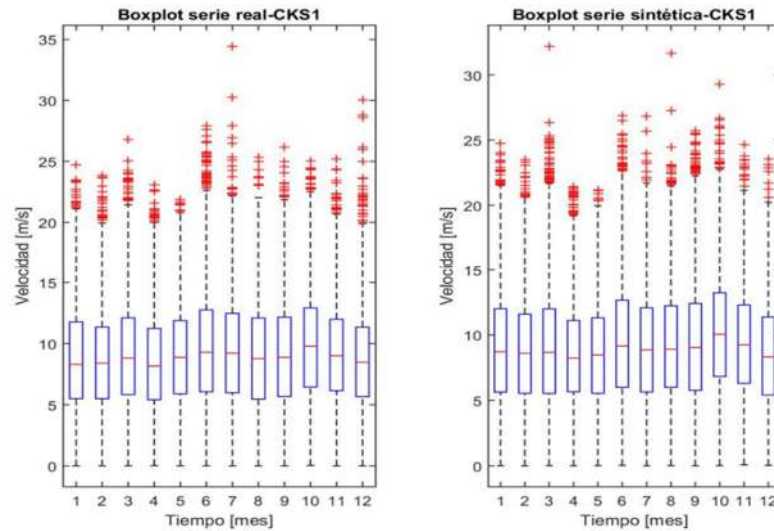


Figura 4.23. Comparación de diagrama de cajas mensuales asociados al sitio CKS1

4.3 Segundo caso de estudio: Planificación de la operación de muy corto plazo

En esta sección se analizará la aplicación de la metodología para un problema ficticio de planificación de la operación de muy corto plazo. En general se tienen los siguientes datos del problema:

- Problema de planificación de la operación de muy corto plazo. Se desean escenarios para un problema de predespacho estocástico.
- El horizonte de evaluación corresponde a 1 día.
- El sistema contiene una central fotovoltaica y otra eólica y se dispone de cerca de un año de mediciones de potencia agregada. Los datos son obtenidos del operador de red del SIC.
- Durante el tiempo de simulación no se prevé la instalación de nuevas centrales de ERV.
- El tiempo de muestreo es de una hora.

- El modelo del sistema eléctrico es simplificado, se consideran las principales líneas de transmisión y se utiliza una simulación cronológica.

Al igual que en el caso anterior se irá aplicando la metodología propuesta etapa por etapa, con la diferencia que, debido a la existencia de la planta fotovoltaica, se dará un mayor énfasis en la incorporación de este recurso en los modelos para generar series sintéticas, mezclando dos metodologías propuestas en la literatura.

Etapas 1: Definir el problema de SEP

Dado los datos entregados es posible contestar las siguientes preguntas de esta etapa:

7. ¿Qué tipo de problema de sistema eléctrico se desea resolver?

Planificación de la operación de muy corto plazo

8. ¿Qué tipo de simulación se utilizará?

Simulación cronológica

9. ¿Qué tipo de modelo del sistema eléctrico es considerado?

Sistema multimodal

10. ¿Cuál es el orden del horizonte de evaluación del problema de SEP?

Días

11. ¿Cuál es el tiempo mínimo de simulación?

Hora (considera tiempos t , tal que $1 \text{ min} < t \leq 60 \text{ minutos}$)

12. ¿Cuáles son los requerimientos de los modelos para centrales renovables?

Modelar sólo plantas existentes

Las respuestas se almacenan y vuelven a ser utilizadas en las etapas 5 y 6.

Etapas 2: Definir tipo de serie del recurso solar y/o eólico disponible

Se cuentan con mediciones de potencia fotovoltaica y eólica en los lugares de interés. Las mediciones consideradas para realizar el análisis estadístico y ajustar los modelos corresponden a centrales conectadas en el SIC. Para efectos de simplicidad y mostrar la aplicación de la metodología así como el modelo se realiza una agrupación de las centrales.

Considerando la descripción del problema ya es posible contestar las preguntas de esta etapa:

1. ¿Cómo se generaron las series con registros de los recursos renovables?

Mediciones del recurso

2. ¿Cuál es el recurso del que se tiene información?

Potencia eólica agregada por región en el sistema eléctrico

Potencia fotovoltaica agregada por región en el sistema eléctrico

3. ¿Cuál es la resolución temporal de los registros?

Hora (considera tiempos t , tal que $1 \text{ min} < t \leq 60 \text{ minutos}$)

Las respuestas son almacenadas para ser utilizadas en las etapas 5 y 6.

Etapas 3: ¿Existe información necesaria para desarrollar el estudio?

El problema considera simulación cronológica, cuyo tiempo de simulación es de una hora. Por otro lado, las mediciones también tienen un tiempo de muestreo de una hora, por lo tanto, existe coherencia entre el estudio y las mediciones, y con ello es posible pasar a la siguiente etapa.

Etapa 4: Análisis estadístico de las series del recurso solar y/o eólico

Se aplica el proceso de la Figura 3.14 a las series de tiempo de potencia fotovoltaica y eólica que son con las que se dispone. La información general de ellas se muestra a continuación:

- Tiempo de muestreo: 1 hora
- Longitud de la muestra: 8544 mediciones
- Fecha: desde el 01-01-2015 hasta el 24-12-2015
- Los sitios analizados son: centrales fotovoltaicas en la región de Atacama y centrales eólicas. Las centrales y sus capacidades se muestran en la Tabla 4.4
- Se deben responder las siguientes preguntas en la etapa 4
 - ¿Qué tipo de función de densidad de probabilidades tienen los recursos?
 - ¿Tiene alguna de las series tendencia?
 - ¿Tiene alguna de las series variaciones estacionales?
 - ¿Existe dependencia temporal?
 - ¿Existe dependencia espacial?
 - ¿Son todas las series estacionarias?

Tabla 4.4. Capacidad instalada de centrales solares fotovoltaicas y eólicas (fuente: CDEC-SIC)

Central	Región	Capacidad instalada [MW]
Solar Chañares	Atacama	36
Solar Diego de Almagro	Atacama	28.05
Solar Esperanza	Atacama	2.88
Solar Javiera	Atacama	65
Solar Las Terrazas	Atacama	3
Solar Llano de Llampos	Atacama	101
Solar PV Salvador	Atacama	68
Solar San Andrés	Atacama	50.6
Solar Santa Cecilia	Atacama	2.96
Eólica Canela 1	Coquimbo	11
Eólica Canela 2	Coquimbo	40
Eólica Totoral	Coquimbo	46
Eólica Monte Redondo	Coquimbo	48
Eólica Talinay	Coquimbo	90
Eólica Punta Colorada	Coquimbo	20
Eólica El Arrayán	Coquimbo	115
Eólica Los Cururos	Coquimbo	109.6
Eólica Punta Palmeras	Coquimbo	45
Eólica Talinay Poniente	Coquimbo	66.6

En detalle se contestan las preguntas de esta etapa usando los métodos gráficos y estadísticos propuestos en el capítulo 3 de este trabajo.

¿Qué tipo de función de densidad de probabilidades tienen los recursos?

Ya que el análisis incluye potencia solar fotovoltaica, automáticamente los recursos no son gaussianos y no es necesario aplicar las pruebas de hipótesis para confirmar. A modo de ejemplo se muestra el histograma de las series de tiempo de potencia fotovoltaica de Atacama en la Figura 4.24. Se observa que debido a las horas en que no está presente la radiación solar hay acumulación de muestras para valores cercanos a cero.

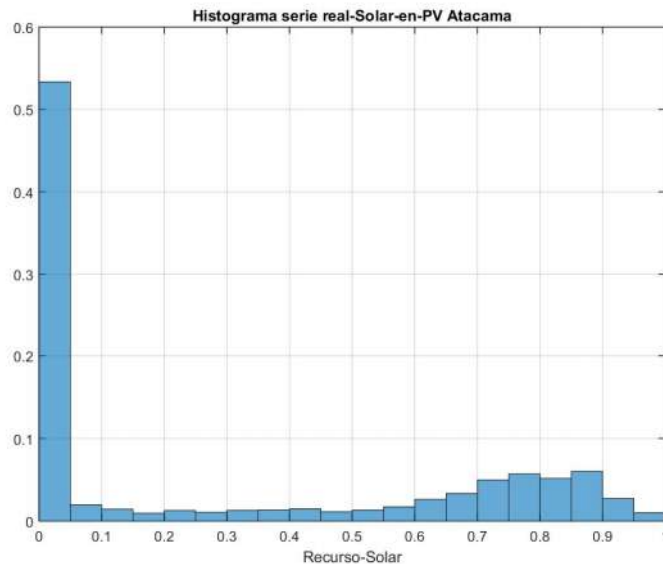


Figura 4.24. Histograma de series de potencia fotovoltaica correspondiente a la región de Atacama (Datos normalizados)

¿Tiene alguna de las series tendencia?

La gráfica en la Figura 4.25 muestra que existe un aumento en la capacidad instalada, por lo tanto las series tienen una tendencia. No es necesario aplicar pruebas de hipótesis para verificar esta condición según fue observado en [104], siendo esta la única metodología presentada en la literatura que lidia con estas situaciones. Para corregir la serie se divide la serie original por los valores máximos a tramos para dejar toda la serie entre cero y uno. Matemáticamente hablando la nueva variable es definida:

$$X'_t = \frac{X_t}{P_t} \quad (3.17)$$

Donde X_t es la serie de tiempo real y P_t es la evolución de capacidad instalada en el tiempo. Se aplicó este cambio de variable porque es una condición para analizar la estacionalidad de las series temporales.

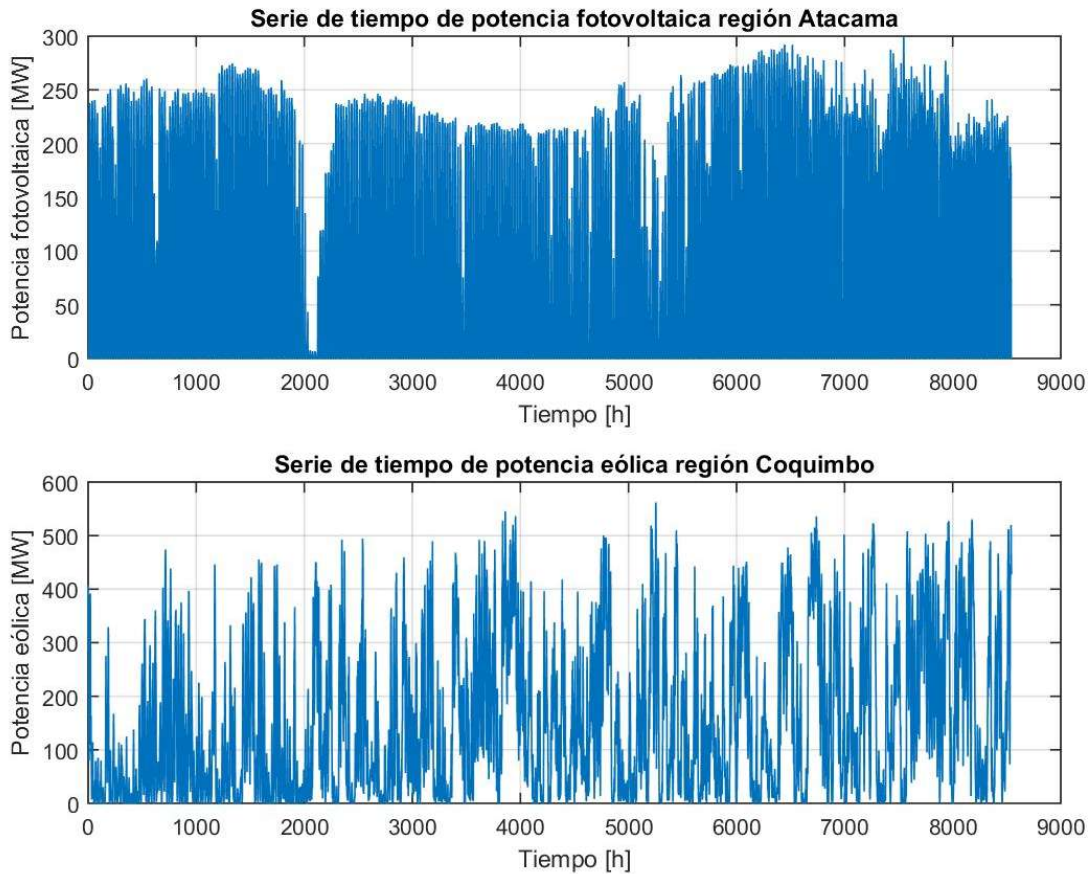


Figura 4.25. Series de tiempo de tiempo de potencia fotovoltaica y eólica en las regiones de Atacama y Coquimbo respectivamente

¿Tiene alguna de las series variaciones estacionales?

Para observar si existen variaciones estacionales, se grafican en la Figura 4.26 los diagramas de cajas para las variaciones horarias. En ambos sitios es posible observar que existe una variación estacional horaria, observándose un patrón notorio en la media durante el día. Esto era de esperar debido a la presencia del recurso solar y no es necesario realizar una prueba estadística. Los diagramas de cajas para este efecto son concluyentes.

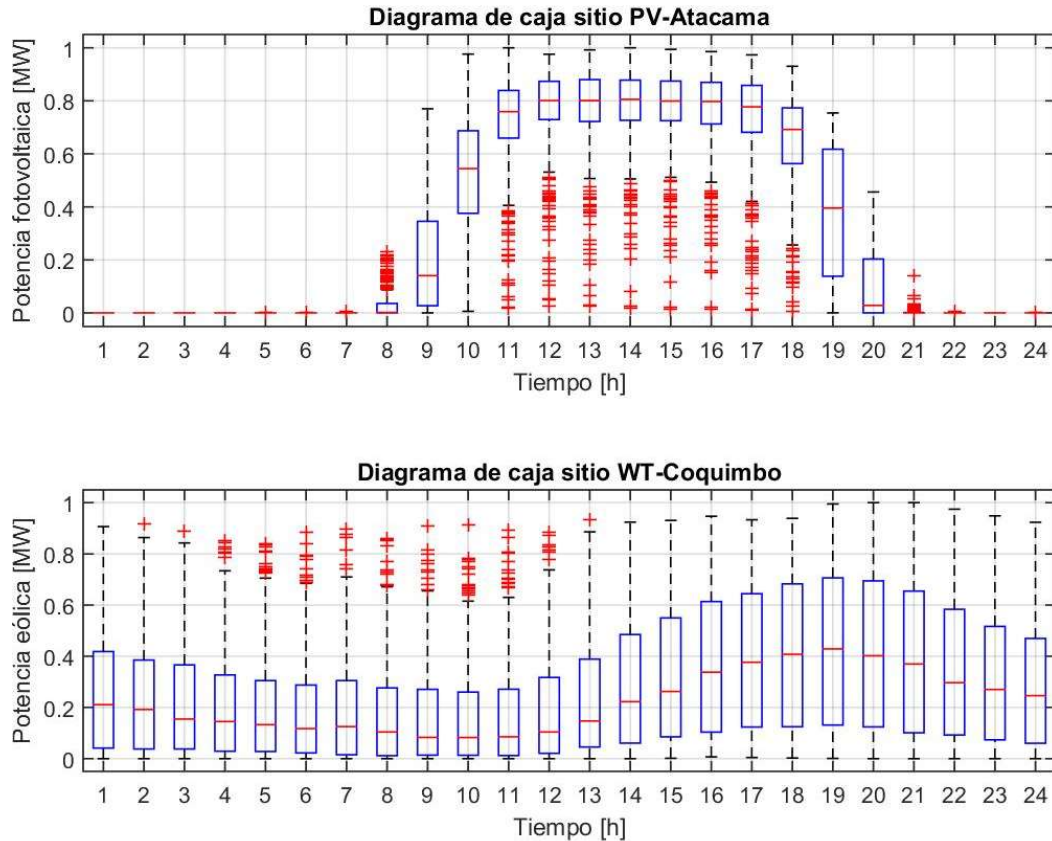


Figura 4.26. Diagrama de cajas para variaciones horarias de series de potencia fotovoltaica y eólica en las regiones de Atacama y Coquimbo

¿Existe dependencia temporal?

La dependencia temporal es observada a través de las funciones de autocorrelación simple (ACF) y parcial (PACF). En la Figura 4.27 se muestran dichas funciones. Dado que existen coeficientes de autocorrelación mayores a esos límites se puede concluir que las series de tiempo se encuentran correlacionadas temporalmente, lo cual se encuentra dentro de lo esperado para estos recursos. Más análisis no son necesarios, con estas funciones es suficiente.

En la ACF asociada a la serie de tiempo de potencia fotovoltaica se observa un peak de correlación cada 12 horas, mostrando la estacionalidad marcada de este recurso. El recurso de potencia eólica tiene una componente estacional horaria al tener un peak en el retardo 24.

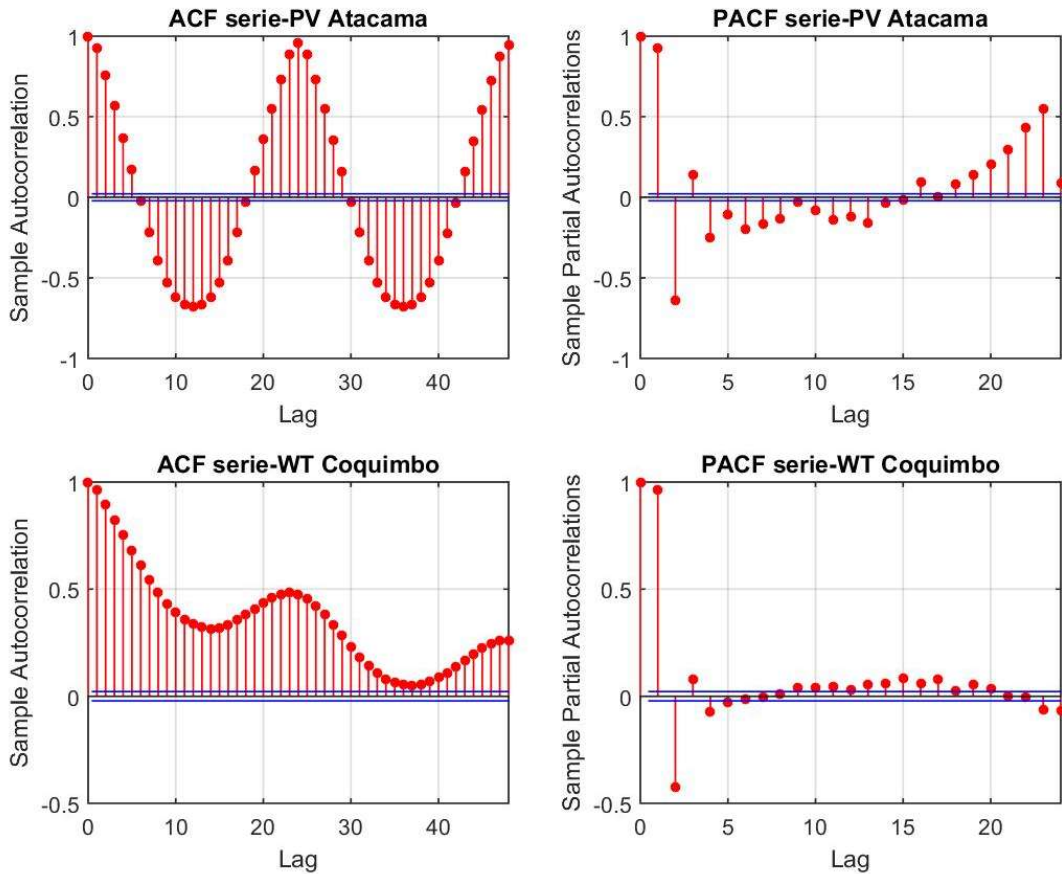


Figura 4.27. Funciones de autocorrelación simple y parcial correspondientes a los sitios PV-Atacama y WT-Coquimbo

¿Existe dependencia espacial?

El análisis de la dependencia espacial se realiza a través de la prueba de hipótesis Ljung Box en su versión multivariable, en especial cuando el número de series de tiempo es mayor a dos. En la Figura 4.28 se muestra que los p-valores son menores al nivel de confianza del 5%, ello implica que se rechaza la hipótesis nula de que la matriz de autocorrelación es cero. Esta conclusión se puede corroborar a través de los gráficos de la función de autocorrelación cruzada y los diagramas de dispersión que se muestran en las Figura 4.29 y Figura 4.30 respectivamente.

Una observación importante debe ser realizada en este análisis. El coeficiente de correlación para el retardo cero es 0.05 y erróneamente se podría decir que no existe correlación significativa entre las variables. Sin embargo, al graficar la función de autocorrelación cruzada (ver Figura 4.30) claramente existen coeficientes mucho mayores a 0.05 para retardos distintos de cero, corroborando la existencia de correlación espacial entre las series que se están estudiando.

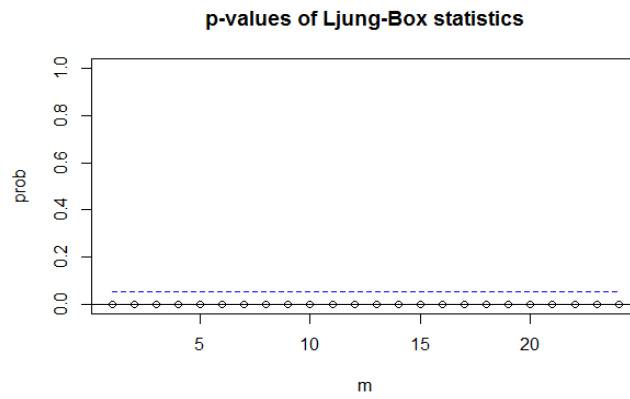


Figura 4.28. Gráfico de p-valores de la prueba de hipótesis Ljung box para modelos multivariados aplicados a las series de potencia fotovoltaica y eólica

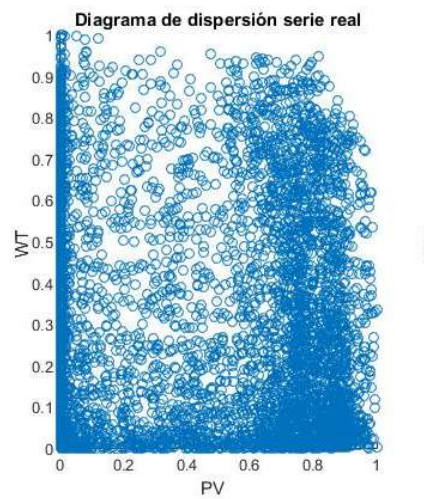


Figura 4.29. Diagrama de dispersión entre sitios PV-Atacama y WT-Coquimbo.

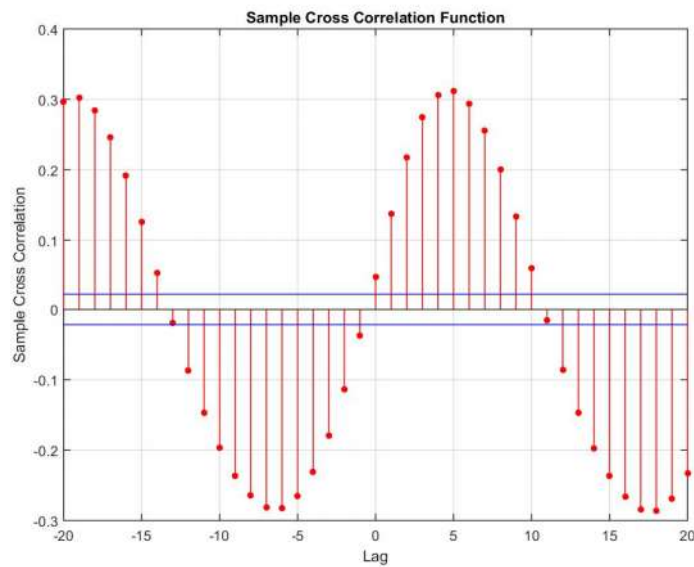


Figura 4.30. Función de autocorrelación cruzada

¿Son las series estacionarias?

Debido a la existencia de estacionalidades las series no son estacionarias. En resumen en esta etapa se tienen las siguientes respuestas al cuestionario:

2. ¿Qué tipo de función de densidad de probabilidades tienen los recursos?

No Gaussiana

2. ¿Tiene alguna de las series tendencia?

Si

3. ¿Tiene alguna de las series variaciones estacionales?

Sí

7. ¿Existe dependencia temporal?

Sí

8. ¿Existe dependencia espacial?

Si

9. ¿Son todas las series estacionarias?

No

Estas respuestas son almacenadas y utilizadas en las etapas siguientes de la metodología propuesta.

Etapas 5: Definir conjunto de modelos candidatos

Siguiendo el proceso descrito en el apartado 3.2.5 y el diagrama de flujo de la Figura 3.17 es posible obtener los modelos o metodologías candidatos. En presencia de series de potencia eólica y fotovoltaica sólo existen dos modelos que pueden ser utilizados. Según la base de datos del Anexo 4 los modelos que corresponden son:

- VAR(p): Corresponde al modelo 19
- Teoría de cópula (diagonal band cópula): corresponde al modelo 28

La aplicación de la familia de diagonal band cópula se encuentra fuera de los alcances de esta tesis. Por lo tanto, se tiene que el modelo a aplicar corresponde a un modelo VAR (p). El modelo 19 propone utilizar la transformada integral por hora. Para eliminar la estacionalidad de corto plazo.

Etapas 6: Selección de modelo

Los datos de entrada a esta etapa son las series de tiempo de mediciones y los modelos candidatos de la etapa anterior. Esta etapa consta de los siguientes procesos:

1. Determinación de procesamiento previo
2. Determinación de parámetros de los modelos
3. Generación de series sintéticas
4. Cálculo de indicadores para evaluar la bondad de los modelos
5. Toma de decisión del modelo

Dado que un modelo candidato corresponde a un proceso VAR, se aplica el diagrama de flujo de la Figura 3.21 para ajustar sus parámetros y determinar el procesamiento y transformaciones necesarias.

Determinación de procesamiento previo

El primer procesamiento consiste en el cambio de variable para estabilizar la media o remover la tendencia. Luego para transformar las series a procesos gaussianos se utiliza la transformada integral por hora (aplica ecuación (3.14)). Con estos procesamientos se tienen series que cumplen con las condiciones necesarias. Un aspecto importante radica en la inclusión de ruido blanco en las horas en las cuales no existe radiación solar; esto fue comentado y agregado en el algoritmo del anexo 3, sección 9.2.

Determinación de parámetros de los modelos

Los parámetros del modelo VAR se realizan a través del criterio AIC y arroja que el modelo debe ser de orden 72. En la Tabla 4.5 se muestra el resumen de los valores que toma el criterio AIC para valores de p entre 66 y 74, siendo el menor valor cuando $p=72$, por ende ese es el orden del modelo. Aplicando las pruebas de hipótesis de autocorrelación a los residuos se observa que cumplen con las condiciones del modelo.

Tabla 4.5. AIC para estimación del orden del modelo VAR(p)

Orden	AIC
66	-3.04058391
67	-3.03999541
68	-3.03962877
69	-3.04029986
70	-3.040104
71	-3.04254718
72	-3.05287212
73	-3.05256953
74	-3.05229243

Dado que existe un solo modelo candidato, los indicadores serán aplicados sólo al modelo VAR(p) y será este el modelo escogido. Los resultados del modelo se mostrarán en la siguiente sección

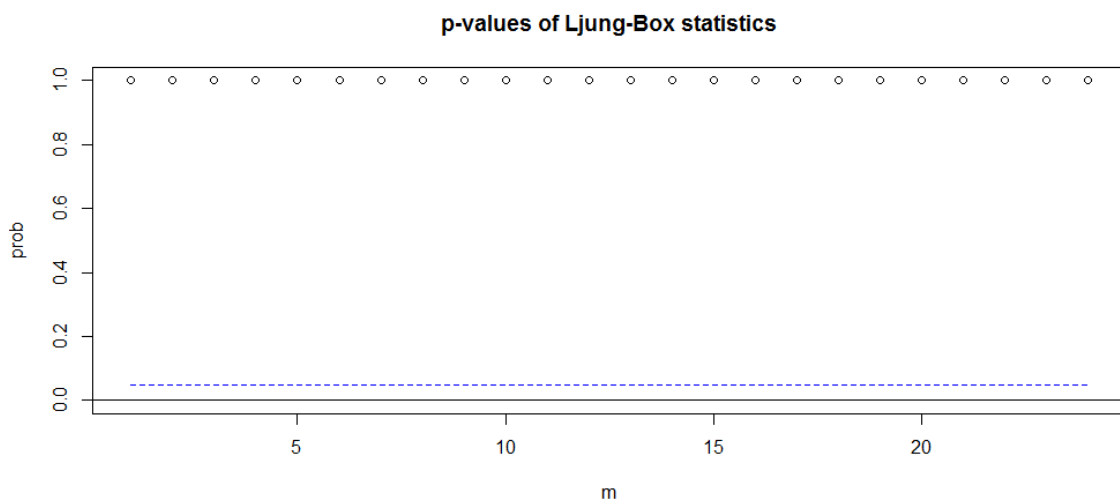


Figura 4.31. Gráfica con aplicación de prueba de hipótesis Ljung-Box para verificar que los residuos no se encuentren correlacionados.

4.3.1 Resultados gráficos del modelo seleccionado

En esta sección se muestran los resultados gráficos y cuantitativos del modelo. En la Figura 4.32 y Figura 4.33 los histogramas entregan información de la función de densidad de probabilidades. Es importante observar que pese a lo poco común de las funciones de densidad que poseen las series de potencia fotovoltaica y eólica, el modelo VAR(p), el cual es lineal, pudo caracterizar adecuadamente esta característica. Las gráficas son muy similares.

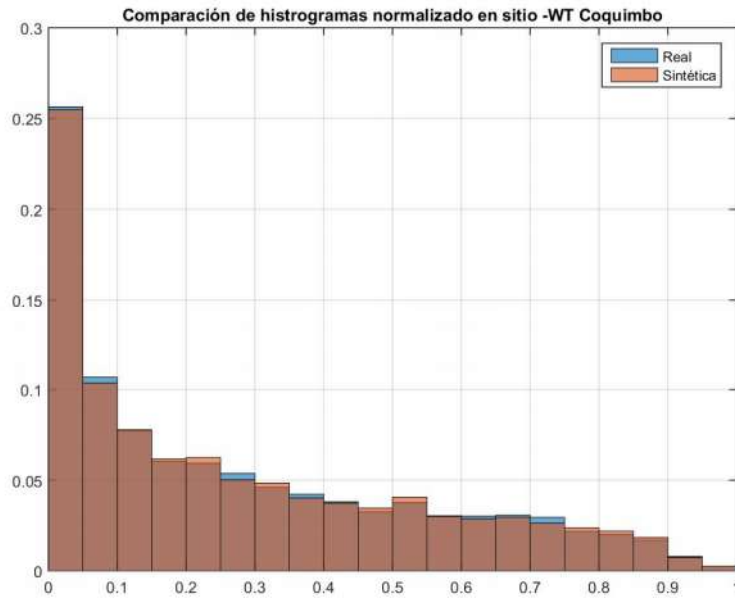


Figura 4.32. Comparación de histogramas en sitio WT-Coquimbo

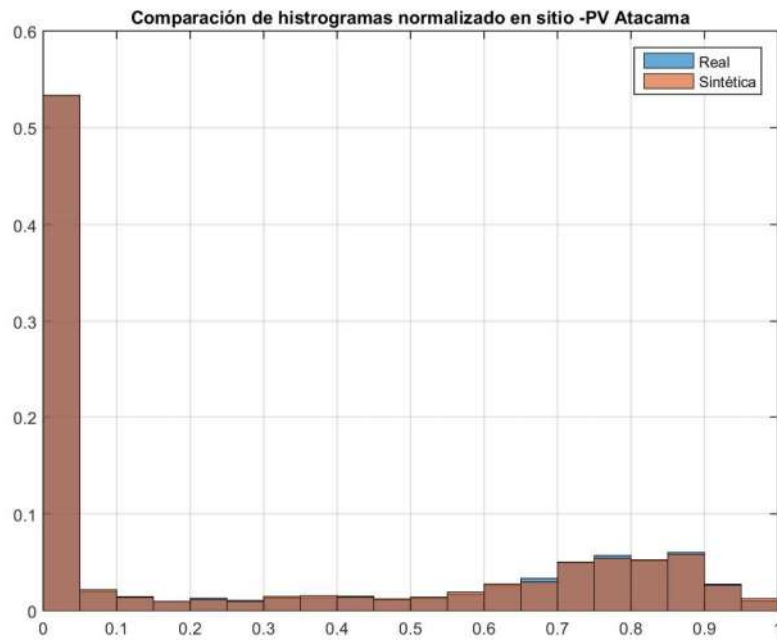


Figura 4.33. Comparación de histogramas en sitio PV-Atacama

Las variaciones estacionales horarias son bien reproducidas por el modelo según se muestran en la Figura 4.34 y Figura 4.35. Tanto la forma como la dispersión son bien caracterizadas. Las ACF, PACF y CCF son muy similares a sus respectivas funciones de las series reales según se observa en la Figura 4.37, Figura 4.38 y Figura 4.39. Un énfasis se debe realizar y es que a pesar que estas funciones son similares a las reales también lo es su estructura de dependencia que se observa a través del diagrama de dispersión de la Figura 4.39. Es un hecho importante, porque a través de un modelo lineal se está caracterizando adecuadamente propiedades que no son lineales.

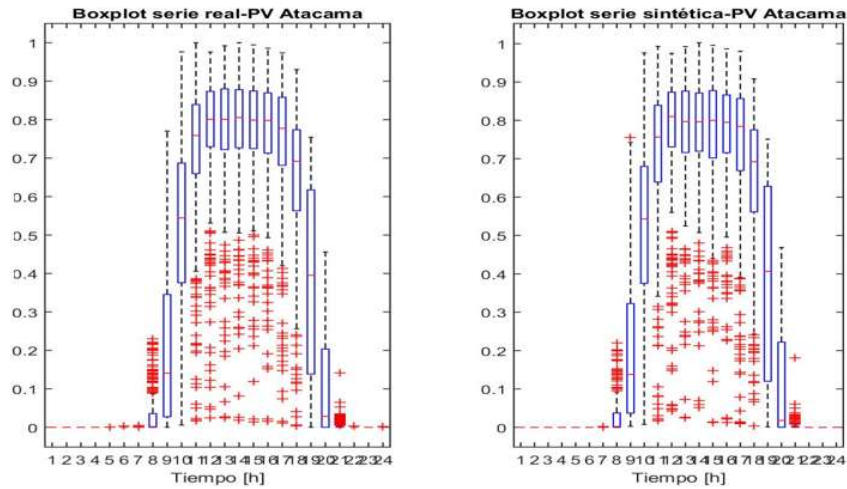


Figura 4.34. Diagrama de cajas en sitio PV-Atacama

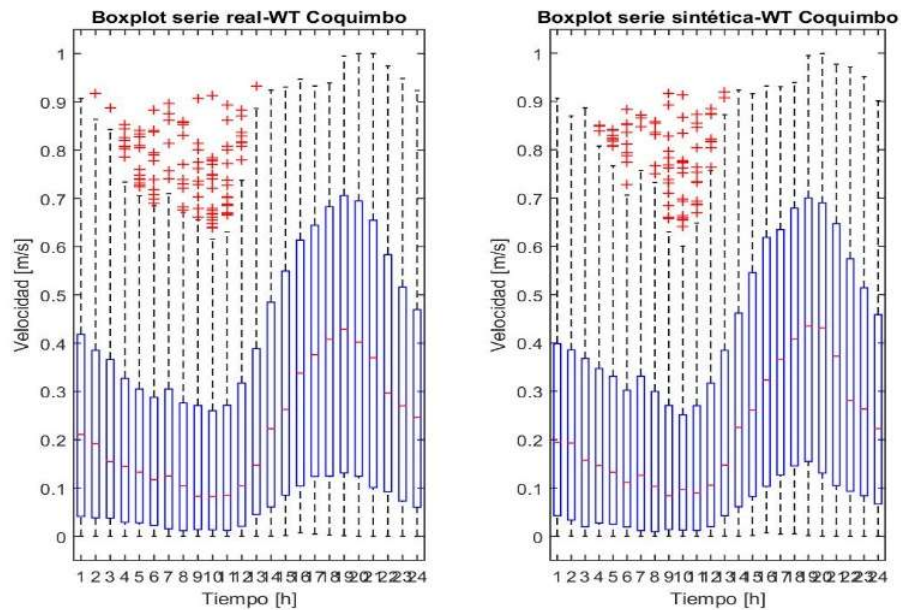


Figura 4.35. Diagrama de cajas para sitio WT-Coquimbo

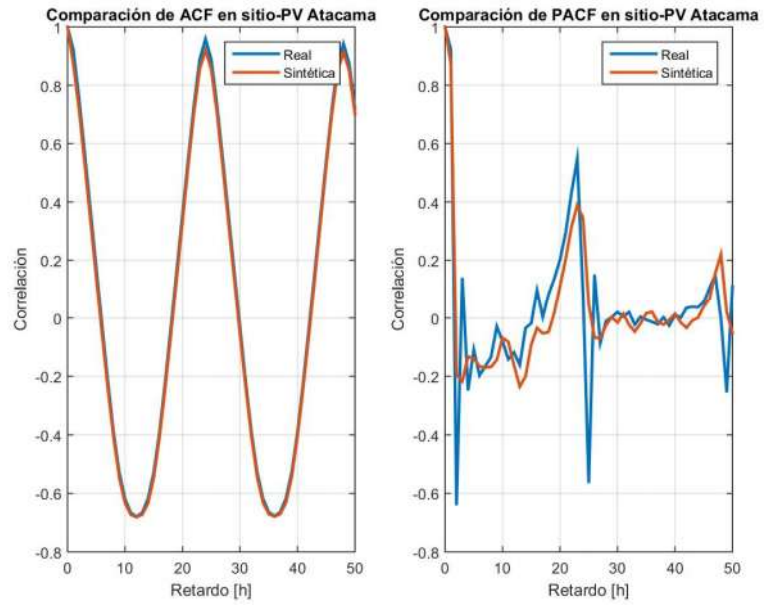


Figura 4.36. Comparación de ACF y PACF en sitio PV-Atacama

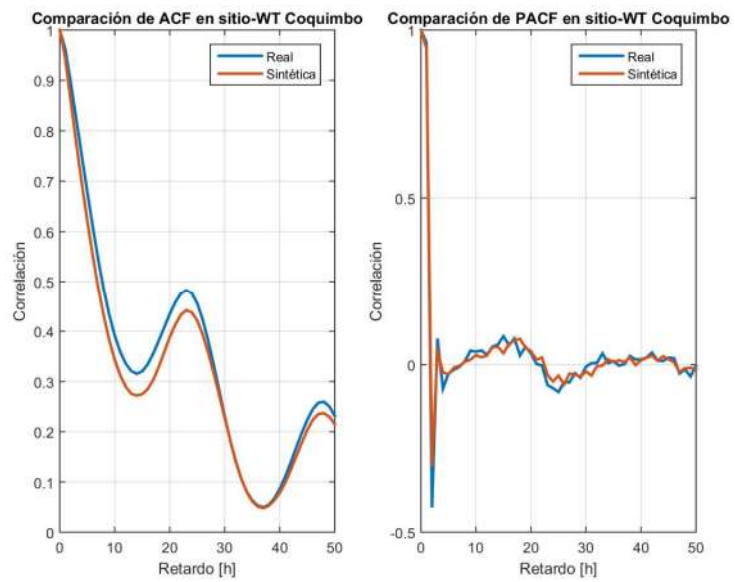


Figura 4.37. Comparación de ACF y PACF en sitio WT-Coquimbo

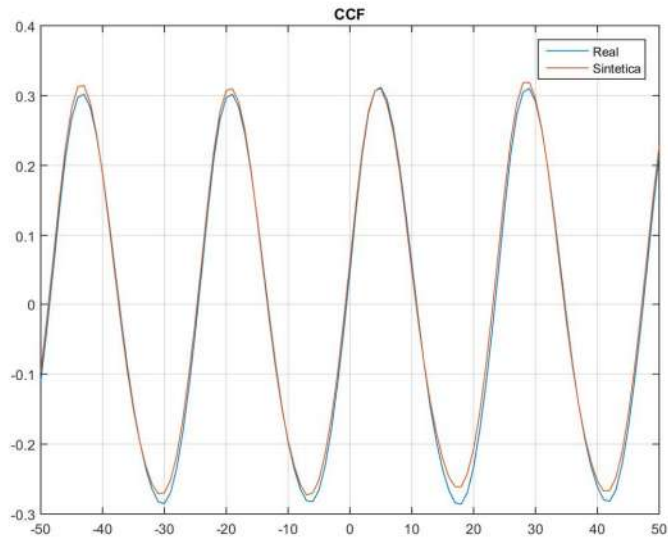


Figura 4.38. Comparación de CCF

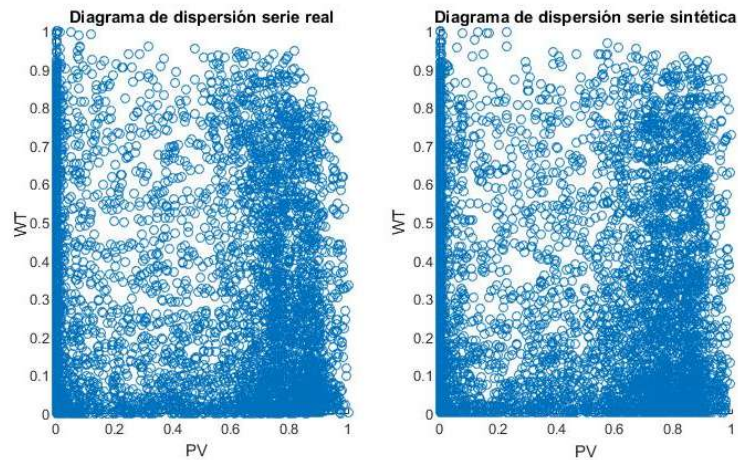


Figura 4.39. Comparación de diagrama de dispersión entre sitio PV-Atacama y WT-Coquimbo

El indicador RMSRE se muestra aplicado para distintos estadísticos en la Figura 4.40. Debido a que el problema que se desea resolver tiene un horizonte de evaluación de un día es importante que las series sintéticas sean capaces de representar las estadísticas horarias. Para este caso, consiste en 24 promedios, desviaciones estándar, cuantil 10%, etc. El indicador aplicado a este conjunto de estadísticos toma un valor menor al 2%, lo que indica que el modelo caracteriza adecuadamente la variabilidad horaria.

El error mayor observa en el RMSRE asociado a la ACF, el cual alcanza un valor del 12% para el sitio WT-Coquimbo y menor al 4% para el sitio PV-Atacama. Dado este resultado se podría concluir que el modelo no está caracterizando bien la dependencia temporal. Sin embargo al analizar la Figura 4.37 efectivamente la serie sintética sí caracteriza adecuadamente la dependencia temporal. Así, un valor RMSRE cerca del 10-12% para la ACF es un valor adecuado y puede utilizarse el modelo sin mayores problemas. Este comentario está basado en la revisión del estado del arte, en donde las gráficas ya expuestas coinciden con distintos resultados pero no se ha aplicado un indicador como RMSRE para analizar el rango de valores permisibles.

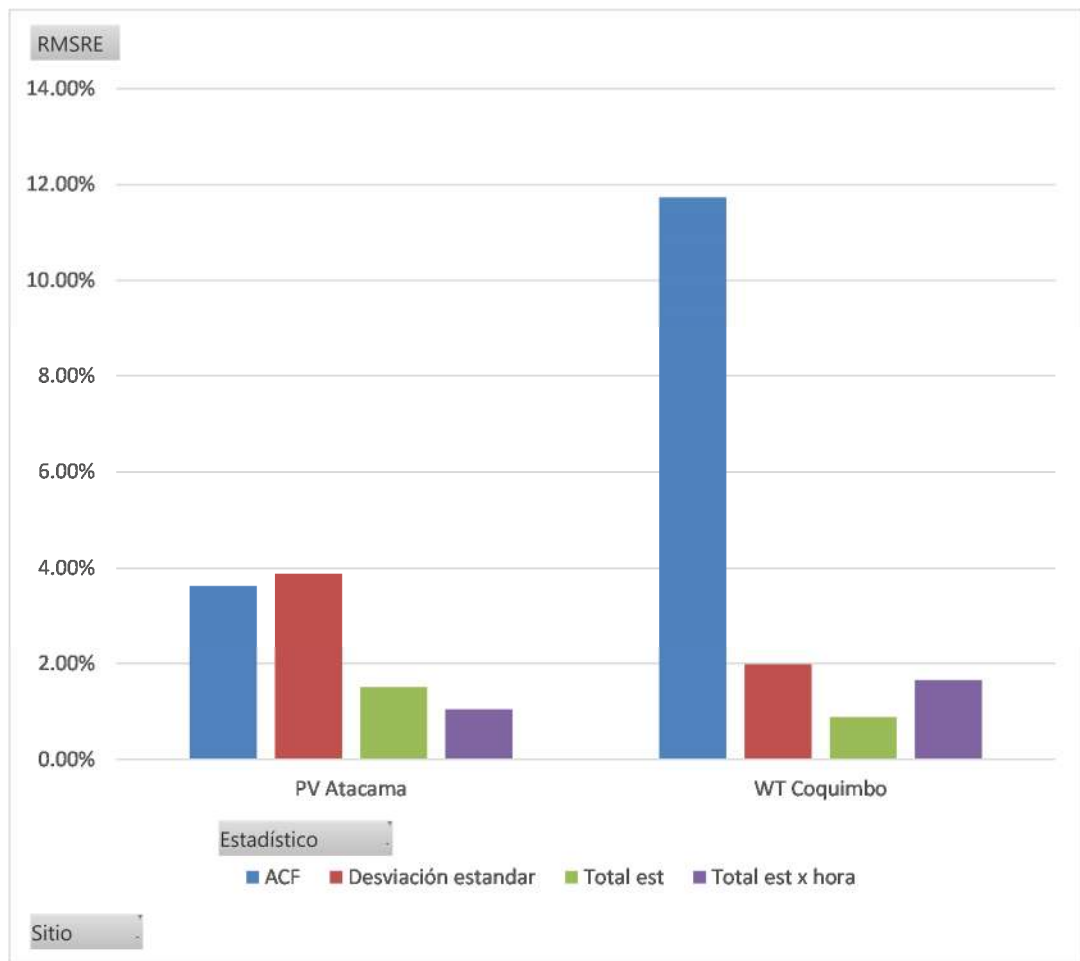


Figura 4.40. RMSRE para distintos estadísticos aplicados al modelo VAR(72)

5 Conclusiones

5.1 Conclusiones respecto al trabajo de tesis

El objetivo principal de esta tesis consiste en conocer los diferentes modelos y/o metodologías para generar series sintéticas de velocidad de viento, potencia eólica, radiación solar y potencia fotovoltaica. Asimismo, a través de dicho estudio, definir criterios que permitan sistematizar el proceso de toma de decisión respecto al modelo a utilizar para caracterizar la variabilidad e incertidumbre de los recursos ERV, todo ello en función de los objetivos de los problemas de SEP que se deseen resolver así como las características estadística propias de los recursos.

Se revisaron las distintas metodologías y modelos utilizados para generar series sintéticas de velocidades de viento, potencia eólica, radiación solar y potencia solar. Además, se revisaron los principales problemas de los SEP con el propósito de detectar el campo de aplicación de las series sintéticas, concluyendo que su aplicación se centra en la planificación de la operación e inversión de SEP. El análisis de los modelos para generar series sintéticas ha sido exhaustivo con el fin de conocer las principales ventajas, desventajas y supuestos de aplicación. El propósito de realizar un análisis de estas características radica en que se busca obtener elementos tales como: criterios, indicadores, procedimientos que permitan jerarquizar, ordenar y sistematizar la selección adecuada de modelos.

Los SEP son sistemas complejos y debido a ello el análisis se divide en horizonte de evaluación y objetivo del estudio. En dicho contexto, son distintas las características que se buscan modelar con mayor detalle en las series sintéticas. Los modelos desarrollados en la literatura son principalmente del tipo estadístico. Los modelos ARMA, ARIMA, teoría de cópula y cadenas de Markov homogéneas y no homogéneas han sido utilizados para generar series sintéticas para los diferentes tipos de ERV que mantengan dependencia temporal. Estos modelos se basan en que las series sean estacionarias, por ello en la literatura se han propuesto diferentes metodologías que permitan transformar series no estacionarias a estacionarias. Con ello se ha logrado caracterizar adecuadamente las estadísticas fundamentales de las series reales. Los modelos para generar series sintéticas que consideren dependencia espacio-temporal son también del tipo estadístico. En especial se han desarrollado modelos VAR, PCA+ARMA y teoría de Cópula. Entre estas estructuras matemáticas, el modelo VAR es el más utilizado debido a que, a través de procesamientos y transformaciones adecuadas, este modelo puede generar buenos resultados en la caracterización estadística de las series de tiempo reales de distinta naturaleza, como lo es la radiación y velocidad de viento.

En base a la revisión bibliográfica se propuso una metodología para sistematizar la selección de modelos. Dicha metodología se basa en seis etapas: definir el problema de SEP, definir el tipo de serie del recurso solar y/o eólico disponible, verificar si existe información necesaria para el desarrollo del estudio, analizar los estadísticos de las series del recurso solar y/o eólico, definir conjunto de modelos candidatos y seleccionar el modelo. Para cada etapa se detalla el procedimiento a seguir, criterios e indicadores, para llegar finalmente a la selección del modelo. La propuesta metodológica fue planteada en forma general, de tal forma de dar flexibilidad en la incorporación de nuevos modelos que puedan ser generados en el futuro.

La metodología es validada a través de dos casos de estudios, considerando las características de problemas de operación de mediano y muy corto plazo. Para ello se desarrollaron los siguientes modelos: cadenas de Markov, ARMA, VAR, CARMA, PCA+ARMA y teoría de cópula Gaussiana, dando un énfasis mayor al modelo VAR por ser el más usado en la literatura. Por todo lo anterior, se propone un proceso específico para sistematizar la selección de transformaciones y/o procesamientos, estimación de orden, verificación de supuesto y por último elección del proceso de muestreo. Los resultados mostraron que la propuesta metodológica ordena y sistematiza el proceso de elección del modelo que caracterice los estadísticos relevantes según el problema de SEP. Se muestra que, el aplicar un modelo o metodología desarrollada en la literatura sin considerar el problema de SEP ni las características propias de las series de tiempo disponibles, puede llevar a utilizar series de tiempo que no modelen características que sean esenciales para las consideraciones que se realizan en los diversos problemas de SEP, tales como: dependencia temporal, estimación de energía según estación, entre otros

5.2 Trabajos futuros

Entre los trabajos futuros que puedan ser realizados con la investigación realizada se encuentran los siguientes:

- La metodología propuesta en este trabajo ordena, jerarquiza y propone indicadores de comparación entre modelos. Con ello se logra sistematizar la selección de modelos. Con ello es posible desarrollar una herramienta computacional basada en un sistema experto que permite automatizar la toma de decisiones para personas que son expertas en SEP pero no en modelos para generar series sintéticas, ayudando a mejorar las prácticas actuales referentes a la caracterización de la variabilidad e incertidumbre de las ERV a través de las series sintéticas.
- En términos de planificación de la operación e inversión, no queda completamente claro los beneficios marginales de utilizar modelos más complejos, o sea, aquellos que puedan ser capaces de caracterizar estructuras de dependencia extremadamente no lineal.
- Existen diversas estructuras entre los modelos multivariados, las cuales aún no han sido aplicadas en el contexto de series sintéticas, tales como: VARMA, VARMA-GARCH, procesos co-integrados.
- En un contexto de estudios que incorporen cientos de plantas renovables, los modelos multivariados se vuelven muy complejos y no se ha estudiado con mayor detalle la aplicación de técnicas de reducción a factores multidimensional para abordar estos problemas.
- Estudiar la factibilidad de incorporar otros criterios para la definición de series sintéticas más allá de los estadísticos relacionados con errores promedio, correlación espacial y temporal. A modo de ejemplo, representación de rampas, ciclos, etc.

6 Bibliografía

- [1] I. B. Andres, “CVAR Constrained planning of renewable generation with consideration of system inertial response, reserve services and demand participation,” Pontificia Universidad Católica de Chile, 2014.
- [2] M. Madrigal and K. Porter, *Operating and planning electricity grids with variable generation*. 2013.
- [3] G. Papaefthymiou, “Integration of stochastic generation in power systems,” Delft University of Technology, 2006.
- [4] A. Flores, “Planificación de la expansión en generación en escenarios de alta penetración de energías renovables,” Tesis de Magister en Ciencias de la Ingeniería, Mención Electricidad, Universidad de Chile, 2014.
- [5] A. J. Conejo, M. Carrión, and J. M. Morales, *Decision Making Under Uncertainty in Electricity Markets*, vol. 153. Boston, MA: Springer US, 2010.
- [6] A. Soroudi and T. Amraee, “Decision making under uncertainty in energy systems: State of the art,” *Renew. Sustain. Energy Rev.*, vol. 28, pp. 376–384, Dec. 2013.
- [7] J. Ekström, M. Koivisto, I. Mellin, J. Millar, E. Saarijärvi, and L. Haarla, “Assessment of large scale wind power generation with new generation locations without measurement data,” *Renew. Energy*, vol. 83, pp. 362–374, 2015.
- [8] P. Chen, “Generation, Stochastic Modeling and Analysis of Power System with Renewable,” Aalborg University, 2010.
- [9] A. Papavasiliou and S. S. Oren, “Stochastic Modeling of Multi-Area Wind Power Production,” *Resources*, vol. 2, pp. 1–10, 2015.
- [10] R. Billinton and W. Wangdee, “Reliability-based transmission reinforcement planning associated with large-scale wind farms,” *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 34–41, 2007.
- [11] H. V. Haghi, S. Member, and S. Lotfifard, “Spatiotemporal Modeling of Wind Generation for Optimal Energy Storage Sizing,” *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 113–121, 2015.
- [12] A. Greenhall, “Wind Scenarios for Stochastic Energy Scheduling,” 2013.
- [13] DNV-GL, “Integration of Renewable Energy in Europe,” 2014.
- [14] IEA, “Key World Energy Statistics 2014,” 2014.
- [15] IEA-ETSAP and IRENA, “Renewable Integration in Power Grids Technology Brief,” 2015.
- [16] IRENA, “Battery Storage for Renewables Market Status and Technology Outlook,” 2015.
- [17] CER, “Centro de Energías Renovables y Ministerio de Energía.” [Online]. Available: <http://cer.gob.cl/blog/2014/10/el-desierto-soleado-de-chile-es-el-mercado-top-de-las-renovable/>. [Accessed: 01-Jan-2015].
- [18] CIFES, “Reporte CIFES- Energías Renovables en el Mercado Eléctrico Chileno,” Santiago, 2015.

- [19] A. Gómez-Expósito, A. Conejo, and C. Cañizares, *Electric energy systems: analysis and operation*. CRC Press, 2009.
- [20] J. Saintcro, “The Effects of Integrating Wind Power on Transimission System Planning, Reliability, and Operations Report on Phase 2: System Performance Evaluation,” 2005.
- [21] N. Tleis, *Power Systems Modelling and Fault Analysis*, vol. 53, no. 9. ELSEVIER, 2013.
- [22] X.-F. Wang, Y. Son, and M. Irving, *Modern Power System Analysis*, vol. 53, no. 9. Springer, 2008.
- [23] A. J. Wood and F. Bruce, *Power Generation, Operation , and Control* .
- [24] B. Carlos, “Herramienta computacional para modelo de predespacho económico de carga,” Universidad de Chile, 2008.
- [25] D. Chattopadhyay, “A practical maintenance scheduling program: mathematical model and case study,” *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1475–1480, 1998.
- [26] J. Carvacho, “Integración de generación eólica en estudios de coordinación hidrotérmica de mediano/Largo Plazo,” Universidad de Chile, 2011.
- [27] R. Hemmati, R.-A. Hooshmand, and A. Khodabakhshian, “Comprehensive review of generation and transmission expansion planning,” *IET Gener. Transm. Distrib.*, vol. 7, no. 9, pp. 955–964, Sep. 2013.
- [28] S. Palmintier, Bryan, “Incorporating operational flexibility into electric generation planning: impacts and metods for system design and policy analysis,” Massachusetts Institute of Technology, 2013.
- [29] J. F. Manwell, J. G. Mcgowan, and A. L. Rogers, *Wind Energy Explained Theory, Design and Application*. Wiley, 2009.
- [30] V. Badescu, *Modeling Solar Radiation at the Earth’s Surface*. Springer, 2008.
- [31] J. Kleissl, *Solar Energy Forecasting and Resource Assessment*, First edit. ELSEVIER, 2013.
- [32] A. Mills, M. Ahlstrom, M. Brower, A. Ellis, R. George, T. Hoff, B. Kroposki, C. Lenox, N. Miller, and J. Stein, “Understanding Variability and Uncertainty of Photovoltaics for Integration with the Electric Power System,” 2009.
- [33] S. Kennedy and P. Rogers, “A Probabilistic Model for Simulating Long-Term Wind-Power Output,” *Wind Eng.*, vol. 27, no. 3, pp. 167–181, 2003.
- [34] D. C. Hill, D. McMillan, K. R. W. Bell, and D. Infield, “Application of auto-regressive models to U.K. wind speed data for power system impact studies,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 134–141, 2012.
- [35] Y. H. Wan, “A primer on Wind Power for Utility Applications,” 2005.
- [36] M. Milligan, K. Porter, E. DeMeo, P. Denholm, B. Kirby, N. Miller, A. Mills, M. O’Malley, and M. Schuerger, “Wind Power Myths Debunked,” *IEEE Power Energy Mag.*, vol. 7, no. 6, pp. 89–99, 2009.
- [37] J. Widén, N. Carpman, V. Castellucci, D. Lingfors, J. Olauson, F. Remouit, M. Bergkvist, M. Grabbe, and R. Waters, “Variability assessment and forecasting of renewables: A review

- for solar, wind, wave and tidal resources,” *Renew. Sustain. Energy Rev.*, vol. 44, pp. 356–375, 2015.
- [38] E. Ela and M. O’Malley, “Studying the variability and uncertainty impacts of variable generation at multiple timescales,” *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1324–1333, 2012.
- [39] J. Ryan, E. Ela, D. Flynn, and M. O’Malley, “Variable generation, reserves, flexibility and policy interactions,” *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, pp. 2426–2434, 2014.
- [40] Y. C. Chen and A. D. Dominguez-Garcia, “A method to study the effect of renewable resource variability on power system dynamics,” *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1978–1989, 2012.
- [41] B. Palmintier and M. Webster, “Impact of unit commitment constraints on generation expansion planning with renewables,” in *IEEE Power and Energy Society General Meeting*, 2011, pp. 1–7.
- [42] N. Troy, E. Denny, and M. O’ Malley, “Base-load cycling on a system with significant wind penetration,” *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 1088–1097, 2010.
- [43] M. Andrés and R. Heresi, “Incorporación de energía eólica en la planificación de la transmisión y operación de sistemas hidrotérmicos,” Pontificia Universidad Católica de Chile, 2008.
- [44] NERC, “Accommodating High Levels of Variable Generation,” 2009.
- [45] A. Shortt, J. Kiviluoma, and M. O’Malley, “Accommodating variability in generation planning,” *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 158–169, 2013.
- [46] A. Sturt and G. Strbac, “Time series modelling of powr output for large scale wind fleets,” *Wind Energy*, vol. 14, no. February 2011, pp. 953–966, 2011.
- [47] J. J. Shynk, *Probability, random variables, and random processes*. Wiley, 2012.
- [48] A. D. Sahin and Z. Sen, “First-order Markov chain approach to wind speed modelling,” *J. Wind Eng. Ind. Aerodyn.*, vol. 89, no. 3–4, pp. 263–269, Mar. 2001.
- [49] H. Nfaoui, H. Essiarab, and a. a. M. Sayigh, “A stochastic Markov chain model for simulating wind speed time series at Tangiers, Morocco,” *Renew. Energy*, vol. 29, no. 8, pp. 1407–1418, Jul. 2004.
- [50] a Shamshad, M. Bawadi, W. Wanhussin, T. Majid, and S. Sanusi, “First and second order Markov chain models for synthetic generation of wind speed time series,” *Energy*, vol. 30, no. 5, pp. 693–708, Apr. 2005.
- [51] F. O. Hocaoglu, O. N. Gerek, and M. Kurban, “The Effect of Markov Chain State Size for Synthetic Wind Speed Generation,” in *Probabilistic Methods Applied to Power Systems, 2008. PMAPS ’08. Proceedings of the 10th International Conference on*, 2008, pp. 2–5.
- [52] E. A. Denaxas, R. Bandyopadhyayi, and D. Patiiio-echeverrrr, “SynTiSe: A Modified Multi-Regime MCMC approach for Generation of Wind Power Synthetic Time Series,” 2015.
- [53] G. Papaefthymiou and B. Kl, “MCMC for Wind Power Simulation,” *IEEE Trans. Energy*

Convers., vol. 23, no. 1, pp. 234–240, 2008.

- [54] K. Brokish, S. Member, and J. Kirtley, “Pitfalls of Modeling Wind Power Using Markov Chains,” in *Power Systems conference and Exposition*, 2009, pp. 1–6.
- [55] T. Wu, S. Member, X. Ai, W. Lin, S. Mem-, J. Wen, and L. Weihua, “Markov Chain Monte Carlo Method for the Modeling of Wind Power Time Series,” pp. 1–6, 2012.
- [56] J. Tang, A. Brouste, and K. L. Tsui, “Some improvements of wind speed Markov chain modeling,” *Renew. Energy*, vol. 81, pp. 52–56, 2015.
- [57] T. Pesch, S. Schröders, H. J. Allelein, and J. F. Hake, “A new Markov-chain-related statistical approach for modelling synthetic wind power time series,” *New J. Phys.*, vol. 17, no. 5, p. 055001, 2015.
- [58] S. Karatepe and K. W. Corcadden, “Wind Speed Estimation: Incorporating Seasonal Data Using Markov Chain Models,” *ISRN Renew. Energy*, vol. 2013, pp. 1–9, 2013.
- [59] R. J. Aguiar, M. Collares-Pereira, and J. P. Conde, “Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices,” *Sol. Energy*, vol. 40, no. 3, pp. 269–279, 1988.
- [60] B. O. Ngoko, H. Sugihara, and T. Funaki, “Synthetic generation of high temporal resolution solar radiation data using Markov models,” *Sol. Energy*, vol. 103, pp. 160–170, May 2014.
- [61] A. Wu, “Expansion Models and Investment Decisions in Electricity Systems with Renewable-induced Uncertainties,” University of Auckland, 2013.
- [62] V. Prema and K. U. Rao, “Development of statistical time series models for solar power prediction,” vol. 83, pp. 100–109, 2015.
- [63] A. Naimo, “A Novel Approach to Generate Synthetic Wind Data,” *Procedia - Soc. Behav. Sci.*, vol. 108, pp. 187–196, 2014.
- [64] J. Jung and R. P. Broadwater, “Current status and future advances for wind speed and power forecasting,” *Renew. Sustain. Energy Rev.*, vol. 31, pp. 762–777, Mar. 2014.
- [65] N. Sansui, A. Zaharim, and K. Sopian, “An Initial Review: Stochastic Application in Wind Speed Forecasting,” *Proceedings, WSEAS 1st Int. ...*, pp. 82–86, 2013.
- [66] A. Sturt and G. Strbac, “Times-series modelling for the aggregate Great Britain wind output circa 2030,” *IET Renew. Power Gener.*, vol. 7, no. 1, pp. 36–44, 2013.
- [67] K. Suomalainen, C. a. Silva, P. Ferrão, and S. Connors, “Synthetic wind speed scenarios including diurnal effects: Implications for wind power dimensioning,” *Energy*, vol. 37, no. 1, pp. 41–50, 2012.
- [68] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, “ARIMA-based time series model of stochastic wind power generation,” *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 667–676, 2010.
- [69] K. Philippopoulos and D. Deligiorgi, “Statistical simulation of wind speed in Athens, Greece based on Weibull and ARMA models,” *Int. J. Energy Environ.*, vol. 3, no. 4, pp. 151–158, 2009.

- [70] J. R. MacCormack, D. Westwick, H. Zareipour, and W. D. Rosehart, "Stochastic modeling of future wind generation scenarios," *40th North Am. Power Symp. NAPS2008*, pp. 1–7, 2008.
- [71] J. L. Torres, a. García, M. De Blas, and a. De Francisco, "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)," *Sol. Energy*, vol. 79, no. 1, pp. 65–77, 2005.
- [72] B. Ernst, B. Kirby, and Y.-H. Wan, "Short-Term Power Fluctuation of Wind Turbines : Analyzing Data from the German 250-MW Measurement Program from the Ancillary Services Viewpoint," *Wind. '99 Conf.*, no. July, pp. 1–12, 1999.
- [73] L. Mora-López and M. Sidrach-de-Cardona, "Multiplicative ARMA models to generate hourly series of global irradiation," *Sol. Energy*, vol. 63, no. 5, pp. 283–291, 1998.
- [74] R. Billinton, H. Chen, and R. Ghajar, "Time-series models for reliability evaluation of power systems including wind energy," *Microelectron. Reliab.*, vol. 36, no. 9, pp. 1253–1261, 1996.
- [75] J. Boland, "Time-series analysis of climatic variables," *Sol. Energy*, vol. 55, no. 5, pp. 377–388, 1995.
- [76] R. Aguiar and M. Collares-Pereira, "TAG: A time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation," *Sol. Energy*, vol. 49, no. 3, pp. 167–174, 1992.
- [77] R. H. Kirchhoff, "A Comparison of Alternative Approaches for the Synthetic Generation of a Wind Speed Time Series," vol. 113, no. November 1991, 1991.
- [78] V. a. Graham, K. G. T. Hollands, and T. E. Unny, "A time series model for K_t with application to global synthetic weather generation," *Sol. Energy*, vol. 40, no. 2, pp. 83–92, 1988.
- [79] B. G. Brown, "Time series models to simulate and forecast wind speed and wind power," *J. Clim. Appl. Meteorol.*, vol. 23, pp. 1184–1195, 1984.
- [80] T. H. E. Activation, E. For, S. Diffusion, and I. N. Ferromagnetic, "Stochastic modeling and forecasting of solar radiation data," *J. Phys. Chem. Solids*, vol. 44, no. 2, pp. 171–174, 1982.
- [81] C. Mustacchi, V. Cena, M. Rocchi, C. Applicata, U. Roma, and R. Italia, "Stochastic Simulation of hourly Global Radiation Sequences," vol. 23, no. 2, pp. 47–51, 1979.
- [82] M. Raby, S. Ríos, S. Jerardino, J. C. Araneda, S. Member, and R. Raineri, "Hydrothermal System Operation and Transmission Planning Considering Large Wind Farm Connection," in *Power tech, 2009 IEEE Bucharest*, 2009, pp. 1–8.
- [83] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Times series analysis forecasting and control*, Third edit. Prentice-Hall, 1994.
- [84] J. M. Morales, R. Mínguez, and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Appl. Energy*, vol. 87, no. 3, pp. 843–855, 2010.
- [85] A. Lojowska, D. Kurowicka, G. Papaefthymiou, and L. Van Der Sluis, "Advantages of ARMA-GARCH wind speed time series modeling," *2010 IEEE 11th Int. Conf. Probabilistic Methods Appl. to Power Syst. PMAPS 2010*, pp. 83–88, 2010.

- [86] B. Klöckl and G. Papaefthymiou, “Multivariate time series models for studies on stochastic generators in power systems,” *Electr. Power Syst. Res.*, vol. 80, no. 3, pp. 265–276, 2010.
- [87] H. Liu, J. Shi, and X. Qu, “Empirical investigation on using wind speed volatility to estimate the operation probability and power output of wind turbines,” *Energy Convers. Manag.*, vol. 67, pp. 8–17, 2013.
- [88] N. Masseran, “Modeling the fluctuations of wind speed data by considering their mean and volatility effects,” *Renew. Sustain. Energy Rev.*, vol. 54, pp. 777–784, 2016.
- [89] Z. Qin, W. Li, and X. Xiong, “Generation system reliability evaluation incorporating correlations of wind speeds with different distributions,” *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 551–558, 2013.
- [90] D. Cai and J. Chen, “Impact of wind speed correlation on planning and operation of distribution network,” no. 0185, pp. 1–4, 2012.
- [91] Y. Li, W. Li, and K. Xie, “Modelling wind speed dependence in system reliability assessment using copulas,” *IET Renew. Power Gener.*, vol. 6, no. 6, pp. 392–399, 2012.
- [92] G. Papaefthymiou and D. Kurowicka, “Using copulas for modeling stochastic dependence in power system uncertainty analysis,” *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 40–49, 2009.
- [93] A. E. Feijóo, J. Cidras, and J. L. G. Dornelas, “Wind speed simulation in wind farms for steady-state security assessment of electrical power systems,” *IEEE Trans. Energy Convers.*, vol. 14, no. 4, pp. 1582–1588, 1999.
- [94] C. Genest and A.-C. Favre, “Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask,” *J. Hydrol. Eng.*, vol. 12, no. 4, pp. 347–368, 2007.
- [95] H. V. Haghi, M. T. Bina, and M. A. Golkar, “Nonlinear Modeling of Temporal Wind Power Variations,” *IEEE Trans. Sustain. Energy*, pp. 1–11, 2013.
- [96] L. Hontoria, J. Aguilera, and P. Zufiria, “Generation of hourly irradiation synthetic series using the neural network multilayer perceptron,” *Sol. Energy*, vol. 72, no. 5, pp. 441–446, 2002.
- [97] R. Carapellucci and L. Giordano, “A new approach for synthetically generating wind speeds: A comparison with the Markov chains method,” *Energy*, vol. 49, pp. 298–305, 2013.
- [98] R. Carapellucci and L. Giordano, “A methodology for the synthetic generation of hourly wind speed time series based on some known aggregate input data,” *Appl. Energy*, vol. 101, pp. 541–550, 2013.
- [99] F. Olsina, “Simulation of Wind Velocity for Stochastic Power Market and Power System Analysis,” *INGENERARE*, no. 27/2013, pp. 67–78, 2013.
- [100] M. Koivisto, “Statistical Analysis of Large Scale Wind Power Generation Using Monte Carlo Simulations,” 2014.
- [101] P. F. Correia and J. M. Ferreira de Jesus, “Simulation of correlated wind speed and power variates in wind parks,” *Electr. Power Syst. Res.*, vol. 80, no. 5, pp. 592–598, 2010.
- [102] B. Klockl, “Multivariate time series models applied to the assessment of energy storage in

- power systems,” ... *Methods Appl. to Power Syst. 2008. PMAPS’08 ...*, 2008.
- [103] M. S. Miranda and R. W. Dunn, “Spatially correlated wind speed modelling for generation adequacy studies in the UK,” *2007 IEEE Power Eng. Soc. Gen. Meet. PES*, 2007.
- [104] M. Koivisto, J. Ekstrom, J. Seppanen, I. Mellin, J. Millar, and L. Haarla, “A statistical model for comparing future wind power scenarios with varying geographical distribution of installed generation capacity,” *Wind Energy*, 2015.
- [105] M. Koivisto, J. Seppänen, I. Mellin, J. Ekström, J. Millar, and I. Mammarella, “Wind Speed Modeling Using a Vector Autoregressive Process with a Time-dependent Intercept Term,” *Electr. Power Energy Syst.*, vol. 77, pp. 1–10, 2016.
- [106] F. Camacho, a. I. McLeod, and K. W. Hipel, “Multivariate contemporaneous ARMA model with hydrological applications,” *Stoch. Hydrol. Hydraul.*, vol. 1, no. 2, pp. 141–154, 1987.
- [107] D. Villanueva, A. Feijóo, and J. L. Pazos, “Simulation of correlated wind speed data for economic dispatch evaluation,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 142–149, 2012.
- [108] J. Nanjing, “Correlation Modeling among Multi-Wind Farms Based on Copula-Arma Wind Speed Model,” in *China International Conference on Electricity Distribution*, 2014, no. Ciced, pp. 23–26.
- [109] A. Y. Abdelaziz, Y. G. Hegazi, M. M. Othman, and W. El-Khattam, “Approach for modelling stochastically dependent renewable energy-based generators using diagonal band copula,” *IET Renew. Power Gener.*, pp. 1–12, 2015.
- [110] D. D. Le, G. Gross, L. Fellow, and A. Berizzi, “Probabilistic Modeling of Multisite Wind Farm Production for Scenario-Based Applications,” *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 748–758, 2015.
- [111] T. Pesch, S. Schröders, H. J. Allelein, and J. F. Hake, “A new Markov-chain-related statistical approach for modelling synthetic wind power time series,” *New J. Phys.*, vol. 17, no. 5, p. 055001, 2015.
- [112] G. D’Amico, F. Petroni, and F. Prattico, “First and second order semi-Markov chains for wind speed modeling,” *Phys. A Stat. Mech. its Appl.*, vol. 392, no. 5, pp. 1194–1201, Mar. 2013.
- [113] S. Kennedy and P. Rogers, “A Probabilistic Model for Simulating Long-Term Wind-Power Output,” *Wind Eng.*, vol. 27, pp. 167–181, 2003.
- [114] B. Klöckl and G. Papaefthymiou, “Multivariate time series models for studies on stochastic generators in power systems,” *Electr. Power Syst. Res.*, vol. 80, no. 3, pp. 265–276, 2010.
- [115] B. Klockl, “Multivariate time series models applied to the assessment of energy storage in power systems,” ... *Methods Appl. to Power Syst. 2008. PMAPS’08 ...*, 2008.
- [116] Y. Li, K. Xie, and B. Hu, “Copula-ARMA model for multivariate wind speed and its applications in reliability assessment of generating systems,” *J. Electr. Eng. Technol.*, vol. 8, no. 3, pp. 421–427, 2013.
- [117] J. M. Morales, R. Mínguez, and a. J. Conejo, “A methodology to generate statistically

- dependent wind speed scenarios,” *Appl. Energy*, vol. 87, no. 3, pp. 843–855, 2010.
- [118] R. Torres, “Modelo de coordinación hidrotérmica multinodal y multiembalse para sistemas eléctricos longitudinales en el mediano plazo,” Universidad de Chile, 2006.
- [119] J. Dowds, P. Hines, T. Ryan, W. Buchanan, E. Kirby, J. Apt, and P. Jaramillo, “A review of large-scale wind integration studies,” *Renew. Sustain. Energy Rev.*, vol. 49, pp. 768–794, 2015.
- [120] W. L. Martinez and A. R. Martinez, *Exploratory Data Analysis with MATLAB*. CRC Press, 2005.
- [121] E. Bash, *Hydrologic Time Series Analysis Theory and Practice*, vol. 1. Springer, 2012.
- [122] D. Peña, *Análisis de series temporales*. Alianza Editorial, 2010.
- [123] W. W. S. Wei, “Time series analysis univariate and multivariate methods,” *International Journal of Forecasting*, vol. 7, no. 3. Pearson, pp. 389–390, 1991.
- [124] E. S. System, “Stable seasonality test,” 2016. [Online]. Available: <https://ec.europa.eu/eurostat/sa-elearning/take-elearning-course-seasonal-adjustment>.
- [125] K. Suomalainen, G. Pritchard, B. Sharp, Z. Yuan, and G. Zakeri, “Correlation analysis on wind and hydro resources with electricity demand and prices in New Zealand,” *Appl. Energy*, vol. 137, pp. 445–462, 2015.
- [126] R. S. Tsay, *Multivariate time series analysis With R and Financial Applications*. Wiley, 2014.
- [127] M. Koivisto, “Finding Value in Big Data Statistical Analysis of Large Data Sets with Applications in Electric Power Systems,” Aalto University, 2015.
- [128] P. M. Lehtonen, M. S. Tech, J. Matilainen, and M. Koivisto, “Statistical Analysis of Large Scale Wind Power Generation,” Aalto University, 2014.
- [129] D. Hill, D. McMillan, K. Bell, D. Infield, and G. W. Ault, “Application of Statistical Wind Models for System Impacts,” in *Universities Power Engineering Conference (UPEC), 2009 Proceedings of the 44th International*, 2009, pp. 1–5.
- [130] G. Roussas, *A Course in Mathematical Statistics*. 1997.
- [131] Helmut Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Second edi. Springer, 2005.
- [132] E. Saeid, *Handbook of Engineering Hydrology Modeling, Climate Change, and Variability*. CRC Press, 2014.
- [133] P. S. P. Cowpertwait and A. V. Metcalfe, *Introductory time series with R*. Springer, 2009.
- [134] T. Schneider and A. Neumaier, “Algorithm 808: ARfit---a matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models,” *ACM Trans. Math. Softw.*, vol. 27, no. 1, pp. 58–65, 2001.
- [135] “Code and data on wind power correlation,” 2016. [Online]. Available: <https://www.stat.auckland.ac.nz/~geoff/>. [Accessed: 30-Oct-2014].

- [136] Dorota Kurowicha; Roger Cooke, *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley, 2006.
- [137] Y. L. Tong, *The Multivariate Normal Distribution*. 1990.
- [138] B. Deler and B. L. Nelson, “Modeling and generating multivariate time series with arbitrary marginals and autocorrelation structures,” *Proceeding 2001 Winter Simul. Conf. (Cat. No.01CH37304)*, vol. 1, 2001.
- [139] D. D. Le, G. Gross, and A. Berizzi, “Probabilistic Modeling of Multisite Wind Farm Production for Scenario-Based Applications,” *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 748–758, 2015.
- [140] P. Poggi, G. Notton, M. Muselli, and a. Louche, “Stochastic study of hourly total solar radiation in corsica using a Markov model,” *Int. J. Climatol.*, vol. 20, no. 14, pp. 1843–1860, 2000.
- [141] R. J. Aguiar, M. Collares-Pereira, and J. P. Conde, “Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices,” *Sol. Energy*, vol. 40, no. 3, pp. 269–279, 1988.
- [142] B. G. Brown, R. W. Katz, and A. H. Murphy, “Time Series Models to Simulate and Forecast Wind Speeds and Power,” *Journal of Climate and Applied Meteorology*, vol. 23, no. August, pp. 1184–1195, 1984.
- [143] A. Sturt and G. Strbac, “Times-series modelling for the aggregate Great Britain wind output circa 2030,” *IET Renew. Power Gener.*, vol. 7, no. 1, pp. 36–44, 2013.
- [144] A. Papavasiliou and S. S. Oren, “Multiarea Stochastic Unit Commitment for High Wind Penetration in a Transmission Constrained Network,” *Oper. Res.*, vol. 61, no. 3, pp. 578–592, 2013.

7 Anexo 1: Fundamentos matemáticos

7.1 Procesos estocásticos estacionarios[123]

Un proceso estocástico es una familia indexadas de variables aleatorias $X_{\omega t}$, donde ω está relacionado a un espacio muestral y t pertenece al conjunto de índices T . Para un valor fijo t , $X_{\omega t}$ es una variable aleatoria. Para un valor dado de ω , $X_{\omega t}$, es una función de t y es llamada realización del proceso. Así, una serie de tiempo es una realización de un proceso estocástico.

En este trabajo se considera que el conjunto de índices es un conjunto de valores enteros. Considere un conjunto de variables aleatorias $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ que pertenece al proceso estocástico $\{X_{\omega t}: t = 0, \pm 1, \pm 2, \dots\}$. La función de distribución conjunta se define por:

$$F_{X_{t_1}, \dots, X_{t_n}}(x_1, \dots, x_n) = P\{\omega: X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n\} \quad (7.1)$$

Donde $x_i, i = 1, \dots, n$ son números reales. Un proceso se dice de primer orden estacionario en distribución si las funciones de distribuciones marginales son invariantes en el tiempo, esto es:

$$F_{X_{t_1}}(x_1) = F_{X_{t_1+k}}(x_1) \quad (7.2)$$

Para cualquier entero t_1, k , y $t_1 + k$. Será de segundo orden estacionario en distribución:

$$F_{X_{t_1}, X_{t_2}}(x_1, x_2) = F_{X_{t_1+k}, X_{t_2+k}}(x_1, x_2) \quad (7.3)$$

Para cualquier $t_1, t_2, k, t_1 + k$ y $t_2 + k$. Por último, será un proceso estacionario de orden n si:

$$F_{X_{t_1}, \dots, X_{t_n}}(x_1, \dots, x_n) = F_{X_{t_1+k}, \dots, X_{t_n+k}}(x_1, \dots, x_n) \quad (7.4)$$

Para cualquier n-tupla (t_1, \dots, t_n) y k enteros. Un proceso se dice que es estrictamente estacionario si (7.4) es verdad para cualquier valor que tome n .

Conociendo que un proceso estocástico, $X_{\omega t}$ es un conjunto de variables aleatorias definidos sobre un espacio muestral, es usual suprimir la variable ω , quedando X_t o $X(t)$. Se definen las funciones media y varianza del proceso:

$$\mu_t = E(X_t) \quad (7.5)$$

$$\sigma_t^2 = E[(X_t - \mu_t)^2] \quad (7.6)$$

Además, se definen las funciones de autocovarianza y autocorrelación:

$$\gamma(t_1, t_2) = E[(x_{t_1} - \mu_{t_1})(x_{t_2} - \mu_{t_2})] \quad (7.7)$$

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}} \quad (7.8)$$

En la práctica es difícil poder verificar si la serie de tiempo es un proceso estrictamente estacionario, por dicha razón se dice que un proceso es débilmente estacionario si:

$$\begin{aligned}
\mu_t &= \mu = cte \\
\sigma_t^2 &= \sigma^2 = cte \\
\gamma(t, t-k) &= E[(x_t - \mu)(x_{t-k} - \mu)] = \gamma_k \quad k = 0, \pm 1, \pm 2 \\
\rho_k &= \frac{\gamma_k}{\gamma_0}
\end{aligned} \tag{7.9}$$

Cuando se habla en este trabajo respecto a series se refiere a que son débilmente estacionarias. Esta es una práctica común en el análisis de series de tiempo.

7.1 Transformada integral [78]

La transformada integral es utilizada para generar muestras pseudo aleatorias provenientes de cualquier familia de función de densidad de probabilidad y es el centro de la aplicación de la teoría de Cópula. En el contexto de series sintéticas es utilizada para transformar series no gaussianas a gaussianas con media cero y desviación estándar uno. La explicación de la transformada integral en esta sección se realiza en función del trabajo publicado en [78], en donde además de transforma una serie de radiación solar a gaussiana, también se aplica para remover la estacionalidad de la serie. Este mismo principio es aplicado en [86], [102].

Sea f_j la función de distribución acumulada horaria que transforma la variable X_t a una Y_t

$$f_j(X_t) = Y_t \tag{7.10}$$

Por simplicidad se escoge la distribución normal estándar para la variable Y_t , de esta forma ambas distribuciones pueden quedar especificadas. La transformación entre ambas variables debe asegurar que sus distribuciones marginales no serán modificadas.

Sea $g(Y_t)$ la distribución normal estándar y $p(X_t)$ la distribución marginal de X_t . Si u es la distribución acumulada de Y_t , se tiene:

$$u = \int_{-\infty}^{Y_t} g(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y_t} e^{-t^2/2} dt \tag{7.11}$$

$$u = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{Y_t}{\sqrt{2}} \right) \right] \tag{7.12}$$

$$\operatorname{erf}(Y_t) = \frac{2}{\sqrt{\pi}} \int_0^{Y_t} e^{-t^2} dt \tag{7.13}$$

De igual forma se define la función de distribución acumulada de X_t como $v = F_{\bar{X}_t}$. Cuando una variable aleatoria es transformada a través de su función de distribución acumulada se genera una nueva variable aleatoria que se distribuye uniformemente entre (0,1), por lo tanto u y v tienen la misma distribución.

Recordando que la distribución de la variable real y la transformada deben ser iguales:

$$\frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{Y_t}{\sqrt{2}} \right) \right] = F_{\bar{X}_t}(\bar{X}_t) \tag{7.14}$$

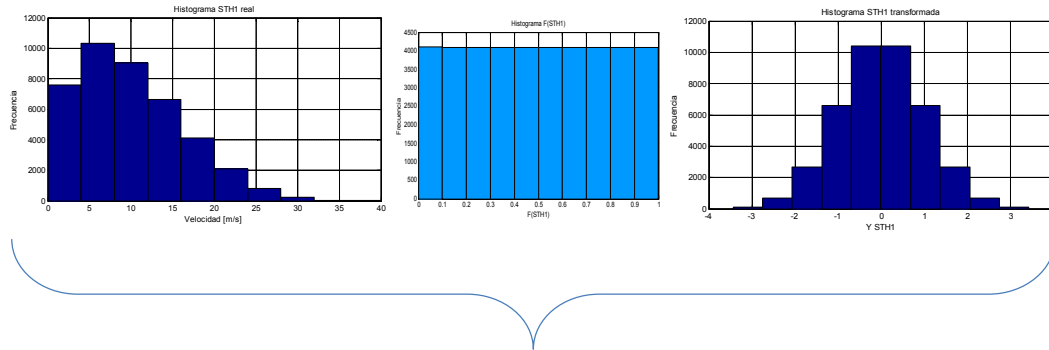
Utilizando la expresión es posible obtener la función que permite realizar la transformación:

$$Y_t = \sqrt{2} \operatorname{erf}^{-1} (2F_{\bar{X}_t}(X_t) - 1) = f_j(X_t) \tag{7.15}$$

El efecto estacional es considerado al obtener f_j para cada hora. Resumiendo, el proceso queda descrito en las siguientes ecuaciones y la Figura 7.1:

$$\vec{Y}_t = F_N^{-1} \left(F_{\vec{X}_t} \left[\vec{X}_t \right] \right) \quad (7.16)$$

$$\vec{X}_t^* = F_{\vec{X}_t}^{-1} \left(F_N \left[\vec{Y}_t \right] \right) \quad (7.17)$$



$$\vec{Y}_t = F_N^{-1} \left(F_{\vec{X}_t} \left[\vec{X}_t \right] \right)$$

Figura 7.1. Aplicación de la transformada integral

Se debe cumplir que las funciones de distribución $F_{\vec{X}_t}$ sean continuas y estrictamente crecientes. Esta condición es usual que se dé con las mediciones de recursos.

7.2 Dependencia entre variables aleatorias

En esta sección se define la independencia entre variables aleatoria para posteriormente definir algunas medidas de correlación que son utilizadas con frecuencia para verificar que también los modelos para generar series sintéticas caracterizan la dependencia temporal y espacial.

7.2.1 Independencia [136]

El concepto de independencia es muy importante en la teoría de probabilidades. Si se tienen dos eventos: A y B, serán independientes si la ocurrencia de uno no influye en el otro. Matemáticamente se puede definir de la siguiente forma:

Definición de independencia: sea un conjunto de variables aleatorias X_1, X_2, \dots, X_n , son independientes si para cada intervalo I_1, I_2, \dots, I_n se cumple [136]:

$$P\{X_1 \in I_1 \cap X_2 \in I_2 \cap \dots \cap X_n \in I_n\} = \prod_{i=1}^n P\{X_i \in I_i\} \quad (7.18)$$

La definición de la ecuación (7.18) es independiente del tipo de variable aleatoria; discreta o continua.

7.2.2 Medidas de dependencia [136][3]

Coefficiente de correlación lineal o Pearson

Este coeficiente es el más utilizado para medir dependencia entre variables aleatorias, para el caso bivariable se tiene la siguiente definición:

El coeficiente de correlación lineal para las variables aleatorias X e Y con esperanzas y varianzas finitas: $E(X)$, $E(Y)$, $\sigma^2(X)$, $\sigma^2(Y)$ es:

$$\rho(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sigma(X)\sigma(Y)} = \frac{E[(X - \mu(X)) - (Y - \mu(Y))]}{\sigma(X)\sigma(Y)} = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)} \quad (7.19)$$

Si se tienen N pares de muestras (x_i, y_i) asociadas a las variables aleatorias (X, Y) es posible estimar el coeficiente de correlación lineal a través de la ecuación:

$$\hat{\rho}_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{\left(\sqrt{\sum_{i=1}^N (x_i - \bar{X})^2} \right) \left(\sqrt{\sum_{i=1}^N (y_i - \bar{Y})^2} \right)} \quad (7.20)$$

Donde:

$$\begin{aligned} \bar{X} &= \frac{1}{N} \sum_{i=1}^N x_i \\ \bar{Y} &= \frac{1}{N} \sum_{i=1}^N y_i \end{aligned} \quad (7.21)$$

El coeficiente de correlación lineal, sólo es definido para aquellas variables aleatorias con esperanzas y varianzas finitas. Si X e Y cumplen con las condiciones descritas en la definición el coeficiente de correlación tiene las siguientes propiedades:

1. Rango:

$$-1 \leq \rho(X, Y) \leq 1 \quad (7.22)$$

2. Independencia: si X, Y son independientes, entonces:

$$\rho(X, Y) = 0 \quad (7.23)$$

3. Invarianza ante transformaciones lineales:

$$\forall a, c \in \mathbb{R}, \rho(aX + b, cY + d) = \text{sgn}(ac)\rho(X, Y) \quad (7.24)$$

4. Dependencia lineal

$$\text{Si } \rho(X, Y) = 1 \Rightarrow \forall a > 0, b \in \mathbb{R}, X = aY + b \quad (7.25)$$

El coeficiente de correlación de Pearson o lineal es una de las medidas más populares para medir la dependencia entre variables aleatorias. Sin embargo, se deben tener algunas consideraciones para interpretar adecuadamente los valores de esta medida. Desde el punto de vista matemático el coeficiente de correlación lineal es solo una medida particular de dependencia entre variables aleatorias. Es una medida canónica en el mundo multivariable normal y en términos más generales para distribuciones esféricas y elípticas. Por lo tanto, cuando se está analizando dependencia entre variables aleatorias que no se distribuyen en forma normal, el coeficiente de Pearson es una sólo una medida de dependencia lineal, esto conlleva a que se pueden despreciar efectos debido a relaciones complejas entre las variables aleatorias.

Coeficiente de correlación Spearman

La idea del coeficiente de correlación de Spearman o Rank correlation consiste en primero ordenar las variables de menor a mayor y luego calcular la correlación lineal entre las variables. Este coeficiente mide la relación monótona entre las variables. De esta forma el efecto de las funciones de distribución es despejada.

El coeficiente de correlación de Spearman de las variables X, Y con funciones de distribución marginal F_X y F_Y es

$$\rho_r(X, Y) = \rho(F_X(X), F_Y(Y)) \quad (7.26)$$

De esta forma el coeficiente de correlación de Spearman es la correlación de Pearson aplicada sobre las variables renqueadas. Este coeficiente posee las siguientes características:

1. Rango:

$$-1 \leq \rho_r(X, Y) \leq 1 \quad (7.27)$$

2. Independencia: si X, Y son independientes, entonces:

$$\rho_r(X, Y) = 0 \quad (7.28)$$

3. Invarianza ante transformaciones no lineales monótonas:

- a. Si $G: \mathbb{R} \rightarrow \mathbb{R}$ es un función estrictamente creciente, entonces:

$$\rho_r(X, Y) = \rho_r(G(X), Y) \quad (7.29)$$

- b. Si $G: \mathbb{R} \rightarrow \mathbb{R}$ es un función estrictamente decreciente, entonces:

$$\rho_r(X, Y) = -\rho_r(G(X), Y) \quad (7.30)$$

4. Dependencia monótona: si $\rho_r(X, Y) = 1$ entonces existe una función estrictamente creciente $G: \mathbb{R} \rightarrow \mathbb{R}, X = G(Y)$

En la aplicación de series sintéticas es este coeficiente el utilizado para medir la correlación entre la serie real y la sintética.

Función de autocorrelación simple [123]

La función de autocorrelación simple (ACF) corresponde a la estimación de los momentos de segundo orden de un proceso estacionario, las cuales están definidas a través del siguiente conjunto de ecuaciones:

$$\gamma_k = \text{cov}[x_t, x_{t+k}] = E[(x_t - \mu)(x_{t+k} - \mu)] \quad (7.31)$$

$$\rho_k = \frac{E[(x_t - \mu)(x_{t+k} - \mu)]}{\sqrt{E[(x_t - \mu)^2]E[(x_{t+k} - \mu)^2]}} \quad (7.32)$$

$$\rho_k = \frac{E[(x_t - \mu)(x_{t+k} - \mu)]}{\sigma_x^2} \quad (7.33)$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (7.34)$$

Estimación de dichos momentos asociados a un proceso estacionario viene dado por:

$$\hat{\mu} = \frac{\sum_{t=1}^T x_t}{T} \quad (7.35)$$

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \hat{\mu})(x_{t+k} - \hat{\mu}) \quad (7.36)$$

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\gamma_0} \quad (7.37)$$

Debido a que corresponde a una estimación, se definen bandas de confianza a través de la ecuación (7.38), en donde si los coeficientes de autocorrelación superan dicha banda, entonces serán significativamente distintos de cero.

$$-\frac{1}{n} \pm \frac{2}{\sqrt{n}} \quad (7.38)$$

Esta función es utilizada como una forma de evaluar si los modelos para generar series sintéticas son capaces de caracterizar la dependencia temporal. Es debido a ello que el objetivo de la serie sintética será emular la ACF asociada a la serie real.

Matriz de covarianza cruzada y correlación [126]

Para medir la dependencia dinámica lineal de un proceso multivariable estacionario x_t se define la matriz covarianza cruzada como:

$$\Gamma_l = Cov(x_t, x_{t-l}) = E[(x_t - \mu)(x_{t-l} - \mu)']$$

$$\Gamma_l = \begin{bmatrix} E(\tilde{x}_{1t}, \tilde{x}_{1,t-l}) E(\tilde{x}_{1t}, \tilde{x}_{2,t-l}) \dots E(\tilde{x}_{1t}, \tilde{x}_{k,t-l}) \\ \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ E(\tilde{x}_{kt}, \tilde{x}_{1,t-l}) E(\tilde{x}_{kt}, \tilde{x}_{2,t-l}) \dots E(\tilde{x}_{kt}, \tilde{x}_{k,t-l}) \end{bmatrix} \quad (7.39)$$

Donde $\mu = E(x_t)$ es el vector de medias de x_t y $\tilde{x}_t = (\tilde{x}_{1t}, \dots, \tilde{x}_{kt}) \equiv x_t - \mu$ es la serie de tiempo centrada en la media. La matriz de covarianza cruzada es una función de l debido a que x_t es un proceso estacionario. Para $l = 0$ se tiene la matriz Γ_0 corresponde a la matriz de covarianza de x_t la cual es común utilizar con la expresión $\Sigma_x = \Gamma_0$.

La matriz de correlación cruzada se obtiene a partir de la ecuación anterior

$$\rho_l = D^{-1} \Gamma_l D^{-1} \quad (7.40)$$

Donde $D = diag\{\sigma_1, \dots, \sigma_k\}$ es una matriz diagonal que contiene las desviaciones estándar de las series dentro de x_t .

Dada una muestra del vector x_t es posible estimar el vector y matriz de covarianza cruzada con la siguiente ecuación:

$$\hat{\mu}_x = \frac{1}{T} \sum_{t=1}^T x_t \quad (7.41)$$

De igual forma es posible estimar la matriz de covarianza y correlación cruzada a través de las ecuaciones (7.42), (7.43) y (7.44)

$$\hat{\Gamma}_0 = \frac{1}{T-1} \sum_{t=1}^T (x_t - \hat{\mu}_x)(x_t - \hat{\mu}_x)' \quad (7.42)$$

$$\hat{\Gamma}_l = \frac{1}{T-1} \sum_{t=l+1}^T (x_t - \hat{\mu}_x)(x_{t-1} - \hat{\mu}_x)' \quad (7.43)$$

$$\hat{\rho}_l = \hat{D}^{-1} \hat{\Gamma}_l \hat{D}^{-1} \quad (7.44)$$

7.3 Muestreo multivariable [137]

Estamos interesados en generar $\vec{\varepsilon}_1, \vec{\varepsilon}_2, \dots, \vec{\varepsilon}_n$ en forma aleatoria y que se distribuya $N(\mu, \Sigma)$, donde μ es fijo y Σ es la matriz de covarianza y es definida positiva. El método depende del siguiente resultado: para cada matriz de orden $n \times n$ y definida positiva $\Sigma = (\sigma_{ij})$ existe una matriz de orden $n \times n$ $T = (\tau_{ij})$ tal que:

$$TT' = \Sigma \quad (7.45)$$

En general la matriz T no es única. Sin embargo, si es restringido el análisis a matrices triangulares inferiores, entonces si es única y puede ser obtenida fácilmente.

Si $\Sigma = (\sigma_{ij})$ es una matriz definida positiva de orden $n \times n$, entonces existe una única matriz triangular $T = (\tau_{ij})$ que satisface $TT' = \Sigma$. Por lo tanto, los elementos de la matriz T vienen dados por:

$$\begin{aligned} \tau_{ij} &= 0 \quad \text{para todo } 1 \leq i < j \leq n \\ \tau_{11} &= \sqrt{\sigma_{11}} \\ \tau_{i1} &= \frac{\sigma_{i1}}{\sqrt{\sigma_{11}}} \quad \text{para } i = 2, \dots, n \\ \tau_{jj} &= \left(\sigma_{jj} - \sum_{r=1}^{j-2} \tau_{jr}^2 \right)^{1/2} \quad \text{para } j = 2, \dots, n \\ \tau_{ij} &= \frac{1}{\tau_{jj}} \left(\sigma_{ij} - \sum_{r=1}^{j-1} \tau_{ir} \tau_{jr} \right) \quad \text{para } j < i \text{ y } i = 2, \dots, n-1 \end{aligned} \quad (7.46)$$

Este resultado es conocido como la factorización de Cholesky

7.4 Criterios AIC

El criterio de información Akaike (AIC por Akaike's Information Criterion) ha mostrado ser de gran efectividad a la hora de estimar el orden de modelos estocásticos estacionarios. Entrega información que relaciona la bondad del modelo y la complejidad de este. Sin embargo, respecto a la bondad no la analiza completamente pero ayuda a iniciar con la estimación del orden de los modelos.

El AIC para modelos ARMA cuyos residuos tienen una varianza estimada σ_e^2 , se obtiene:

$$AIC = 2k + N \ln(\hat{\sigma}_e^2) \quad (7.47)$$

El criterio AIC se extiende para versiones multivariantes quedando la ecuación:

$$AIC(l) = \ln \left| \hat{\Sigma}_{a,l} \right| + \frac{2}{T} lk^2 \quad (7.48)$$

Donde T es el tamaño de la muestra, $\hat{\Sigma}_{a,l}$ es el estimador de la matriz de covarianza y l el orden del modelo $VAR(l)$.

Para estimar el orden de p y q para los modelos ARMA(p, q) y el orden p para los modelos VAR(p) utilizando AIC se debe aplicar el siguiente algoritmo:

1. Definir $p=q=0$
2. Definir p_{\max} y q_{\max}
3. Incrementar p y q en uno
4. Calcular AIC para el orden del paso anterior
5. Se escoge el orden del modelo que tenga el menor AIC

8 Anexo 2: Modelos que consideran dependencia temporal

8.1 Antecedentes generales del anexo

En este anexo se comentan en forma resumida los algoritmos para generar series sintéticas que consideran dependencia temporal. Los modelos escogidos son cadenas de Markov y modelos ARMA debido a las ventajas que se observaron en el estado del arte de este trabajo. Además se muestran resultados de la aplicación de estos algoritmos al sitio CKS1. Para este sitio se considera el recurso eólico y la variable que se analiza es la velocidad de viento. El enfoque de este anexo consiste en la aplicación más que en un análisis matemático riguroso, y es a través de los resultados que se verifica la bondad de dichos modelos y los problemas que pueden ocurrir en la aplicación.

8.2 Cadenas de Markov [50], [51],[53]

Una cadena de Markov es un proceso estocástico, y es de primer orden cuando la probabilidad del estado siguiente depende del estado actual, en efecto sea $X(t)$ un proceso estocástico, el cual posee un espacio de estado finito $S = \{1, 2, \dots, K\}$. En general para una secuencia en el tiempo $t_1 < t_2 < \dots < t_{n-1} < t_n$, la probabilidad condicional viene dada por:

$$P\{X(t_n) = i_n | X(t_1 = i_1), \dots, X(t_{n-1} = i_{n-1})\} = P\{X(t_n) = i_n | X(t_{n-1}) = i_{n-1}\} \quad (8.1)$$

$$P\{X(t) = j | X(s) = i\} = P_{ij}(s, t) \quad (8.2)$$

Si se tienen n estados, habrá $n \times n$ transiciones entre dos instantes sucesivos del tiempo. Por lo tanto, es posible encontrar la probabilidad de transición p_{ij} que va desde el estado t al estado $t+1$.

Se define la matriz de transición de probabilidad $P_{t,t+1}$

$$P_{t,t+1} = \begin{bmatrix} p_{1,1} & p_{1,2} & \dots & p_{1,k} \\ p_{2,1} & p_{2,2} & \dots & p_{2,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k,1} & p_{k,2} & \dots & p_{k,k} \end{bmatrix} \quad (8.3)$$

Los elementos de la matriz de transición se pueden calcular con las frecuencias en las cuales el estado i va hacia el estado j , de esta forma:

$$p_{ij} = \frac{n_{ij}}{\sum_j n_{ij}} \quad (8.4)$$

Análogamente, para el caso de una cadena de segundo orden, la Matriz de probabilidades queda definida:

$$P_{t-1,t,t+1} = \begin{bmatrix} p_{1.1,1} & p_{1.1,2} & \cdots & p_{1.1,k} \\ p_{1.2,1} & p_{1.2,2} & \cdots & p_{1.2,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{1.k,1} & p_{1.k,2} & \cdots & p_{1.k,k} \\ p_{2.1,1} & p_{2.1,2} & \cdots & p_{2.1,k} \\ p_{2.2,1} & p_{2.2,2} & \cdots & p_{2.2,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k.k,1} & p_{k.k,2} & \cdots & p_{k.k,k} \end{bmatrix} \quad (8.5)$$

La probabilidad p_{ijk} es la probabilidad de que en el siguiente estado de velocidad sea k si en el estado actual es j y en el estado previo es i . La forma de calcular la matriz P es análoga a la cadena de primer orden y la suma de cada fila es igual a uno.

Las ecuaciones anteriores definen una cadena de Markov homogénea, esto significa que la matriz de transición de probabilidades P es invariante en el tiempo. Desde una mirada de series de tiempo significa que la serie de tiempo es estacionaria. Las cadenas homogéneas solo pueden ser utilizadas cuando no existen variaciones estacionales muy marcadas. Como principal ventaja tiene que puede modelar procesos cuya función de densidad de probabilidad sean no gaussianas.

A continuación se especifica el algoritmo para generar series sintética utilizando un modelo de cadena de Markov de primer orden. El algoritmo ha sido utilizado en [53].

Algoritmo:

1. Definir cantidad de estados
2. Discretizar en función de la cantidad de estados del paso anterior
3. Obtener la matriz de transición de probabilidades P
4. Calcular la matriz de probabilidades acumulada
5. Iniciar con el primer estado en forma aleatoria
6. Generar un número aleatorio entre 0 y 1 utilizando una distribución uniforme de probabilidad
7. Para definir el siguiente estado, el número aleatorio u_j obtenido en el paso 6 es comparado con la i –ésima fila de la matriz de probabilidad acumulada. Si u_j es mayor que la probabilidad acumulativa del estado previo pero menor o igual a la probabilidad acumulada del estado siguiente, el siguiente estado es adoptado
8. La velocidad del viento o potencia es calculado con la expresión:

$$v_j = v_{j\min} + u_j (v_{j\max} - v_{j\min}) \quad (8.6)$$

Para ayudar a entender la aplicación de este modelo se aplicará al sitio CKS1. En primer lugar se definen los estados en función del histograma de datos. Para este caso y dada las muestras se decide utilizar 70 estados discretos y están espaciados por 0.5 m/s. Una vez discretizada las mediciones se estima la matriz de transición de probabilidades y la matriz de transición de probabilidades acumulada. Con estos datos genera la serie sintética de velocidades de viento.

Los resultados de los estadísticos considerando toda la muestra se presentan en la Tabla 8.1. De allí se observa que el modelo entrega resultados muy similares a los reales. Este hecho también es corroborado a través de la gráfica de los histogramas según se ve en la Figura 8.1

Tabla 8.1. Comparación de estadísticos entre serie real y sintética en sitio CKS1 utilizando un modelo de cadena de Markov de primer orden

N°	Tipo modelo	Promedio	Desviación estándar	Varianza	Mínimo	Máximo	Q10	Q50	Q75	Q90
1	Real	9.057	4.411	19.453	0.000	34.438	3.470	8.803	11.986	14.908
8	MCMC orden 1	9.133	4.492	20.176	0.000	34.030	3.483	8.892	12.031	15.075

Las funciones ACF y PACF tienen un comportamiento de esperar para una cadena de Markov de primer orden, esto implica que experimentan un decaimiento exponencial a medida que aumentan los retardos y no es posible que se modelen los efectos estacionales. En la serie real, la ACF tiene un peak en el retardo 24 indicando que existe un grado de dependencia. Según el análisis estadístico presentado en el apartado 4.2 de esta tesis, la serie posee variaciones estacionales horarias y ello implica que este modelo a pesar de que sea capaz de modelar bien los estadísticos totales, cuando se calculan los estadísticos por hora el error aumenta. Es posible apreciar este efecto en la gráfica del indicador RMSRE que se muestra en la Figura 8.3. Los errores son mayores al 10% para el caso del promedio horario (24 promedios considerando hora =1,...,24) y la ACF llega al 25%.

Los resultados muestran las ventajas de usar este modelo, pero también los cuidados que se deben tener para saber cuándo es posible utilizar cadenas de Markov. Es posible superar las desventajas pero el modelo crece en complejidad, si se desea modelar mejor la ACF se debe aumentar el orden de la cadena, la que crece en forma exponencial, lo que hace más compleja la estimación de la matriz de transición de probabilidades. Para el caso de modelar las estacionalidades se ha propuesto considerar cadenas no homogéneas, entonces se define una cadena por estación, ello también tiene problemas de aumentar la cantidad de parámetros a estimar pero sigue siendo relativamente sencillo el procedimiento y algoritmo para generar series sintéticas.

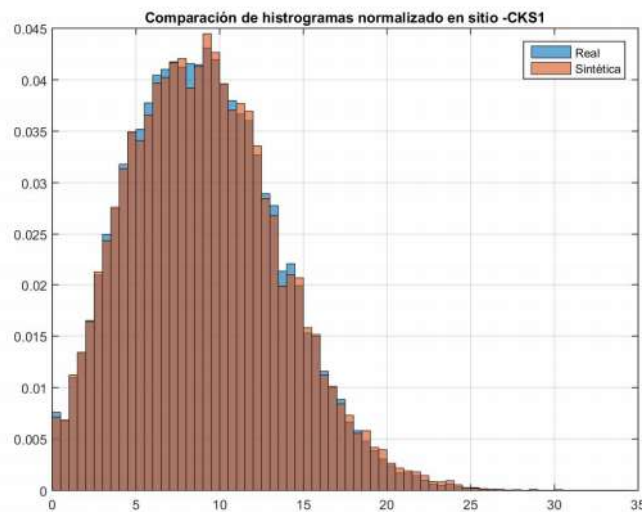


Figura 8.1. Comparación de histogramas en sitio CKS1 utilizando cadena de Markov de primer orden

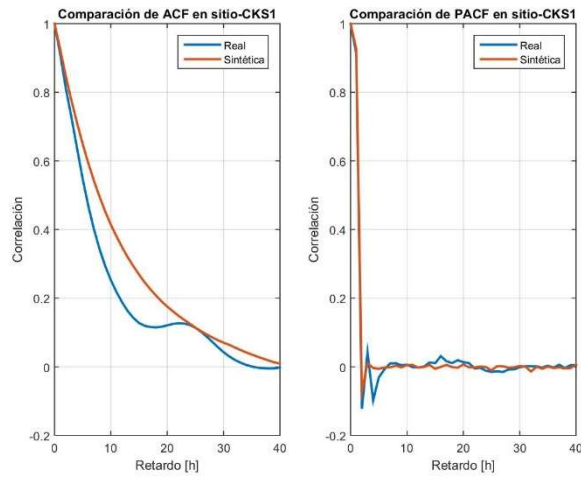


Figura 8.2. Comparación de ACF y PACF en sitio CKS1 utilizando cadena de Markov de primer orden

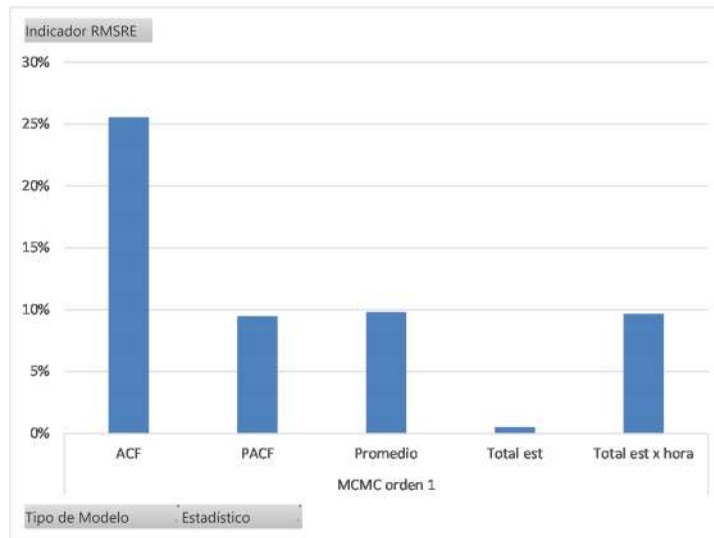


Figura 8.3. Gráfica de indicador pare diferentes estadísticos del sitio CKS1 utilizando cadena de Markov de primer orden

8.3 ARMA

Los modelos ARMA (p,q) han sido desarrollados durante las últimas décadas siendo de gran utilidad para explicar diferentes fenómenos en econometría, series climatológicas entre otras. Los modelos ARMA (p,q) pueden ser descritos mediante la ecuación (8.7), donde los parámetros ϕ_i, θ_j con $i = 1, \dots, p$ y $j = 1, \dots, q$ corresponden a los coeficientes auto regresivos y de media móvil del modelo. ε_t corresponde a un proceso ruido blanco que es independiente e idénticamente distribuido de acuerdo a una distribución normal $N(0, \sigma_\varepsilon)$

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (8.7)$$

Para poder aplicar correctamente este modelo la serie debe ser débilmente estacionaria, de lo contrario es necesario aplicar transformaciones para cumplir con esta condición. Además, para poder representar adecuadamente las funciones de densidad de probabilidades se debe aplicar la transformada integral para

convertir a procesos gaussianos con media cero y varianza uno. Lo anterior es equivalente a lo expuesto en el anexo 3 para los modelos VAR. Los procesamientos y transformaciones utilizadas en la literatura son:

$$Y_t = \frac{X_t - u_{tk}}{\sigma_{tk}}, \quad \forall k \in \{S\} \quad (8.8)$$

$$X'_t = \frac{X_t - u_{tk}}{\sigma_{tk}}, \quad \forall k \in \{S\} \quad (8.9)$$

$$Y_t = F_{invN}^{-1} \left(F \left(X'_t \right) \right)$$

$$Y_t = F_{invN,k}^{-1} \left(F \left(X_t \right) \right), \quad \forall k \in \{S\} \quad (8.10)$$

Una vez que se logra cumplir con las condiciones impuestas por el modelo, éste se debe ajustar y verificar las hipótesis que se realizan sobre los residuos.

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (8.11)$$

En las ecuaciones anteriores, F_{invN}^{-1} es la función de distribución acumulada de probabilidad normal estándar, F es la función de distribución acumulada de probabilidad de la serie de tiempo X'_t , X'_t es la serie sin la componente estacional y por último Y_t es la serie transformada a la cual se ajusta el modelo.

Finalmente el algoritmo para generar series sintéticas es el siguiente:

1. Remover estacionalidades a través de las ecuaciones (8.8), (8.9)
2. Aplicar transformada integral
3. Estimar orden del modelo utilizando criterio AIC
4. Verificar hipótesis sobre los residuos (autocorrelación y normalidad) Si no se cumple, aumentar el orden p del modelo hasta que las hipótesis sobre los residuos se satisfagan
5. Generar serie sintéticas
6. Aplicar transformaciones inversas para volver al espacio real

A continuación se muestra un ejemplo de la aplicación de este modelo para generar series sintéticas. Se utiliza las mediciones del mismo sitio usado para el modelo de cadenas de Markov, CKS1.

Se aplica la ecuación (8.8) para remover estacionalidad horaria y luego la transformada integral para convertir a un proceso gaussiano. Luego se ajusta el modelo ARMA(3,3) utilizando el criterio AIC, resultando con ello que las hipótesis sobre los residuos se cumplen. En la Figura 8.4 se muestra la ACF y PACF de los residuos, en donde se observa que los coeficientes de autocorrelación son menores a la banda de significancia indicando que los residuos no se encuentran correlacionados.

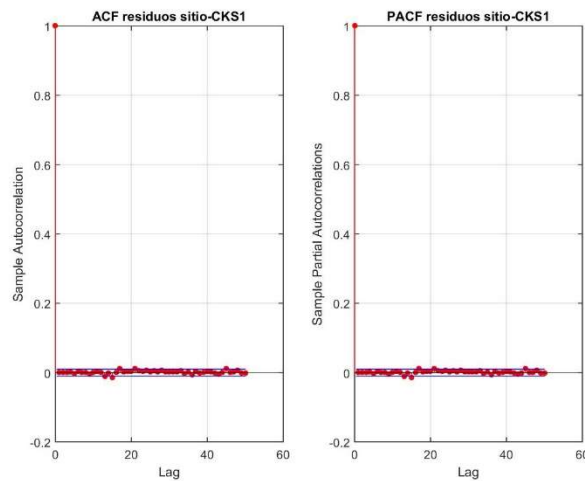


Figura 8.4. ACF y PACF de los residuos del modelo ARMA(3,3) asociado al sitio CKS1

En las ACF y PACF que se muestran en la Figura 8.4 es posible observar que existe gran similitud entre la serie real y la serie sintética. Por otro lado la función de densidad de probabilidades así como las variaciones estacionales horarias también son similares entre ambas series. Los indicadores de la Figura 8.8 ratifican el análisis realizado a los gráficos y además se observa un RMSRE del 10% para la ACF. Se podría decir que es grande el error, pero de acuerdo a la Figura 8.3 sí son similares, debido a ello un 10% RMSRE en la ACF no es un valor elevado. Además con este ejemplo se muestra que, a través de transformaciones, es posible caracterizar funciones de densidad de probabilidades distintas a la gaussiana con los modelos ARMA, recordando que es la principal desventaja que se exponía en la literatura especializada.

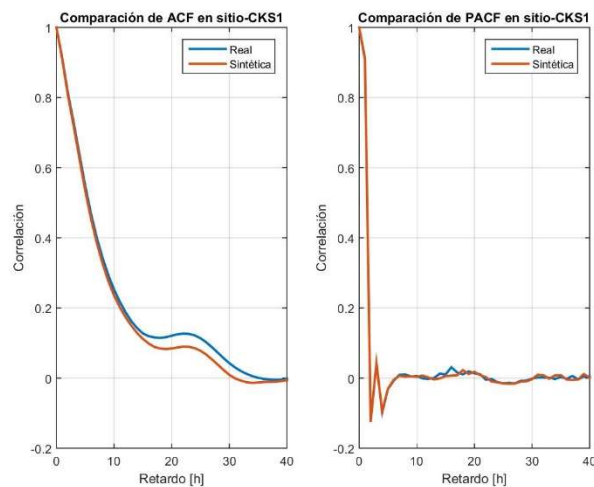


Figura 8.5. Comparación de ACF y PACF en sitio CKS1 utilizando modelo ARMA(3,3)

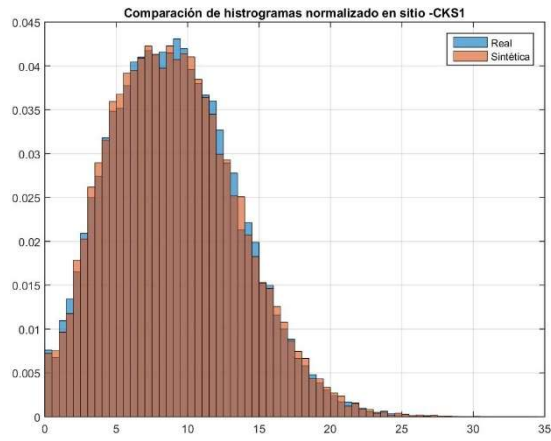


Figura 8.6. Comparación de histogramas en sitio CKS1 utilizando modelo ARMA(3,3)

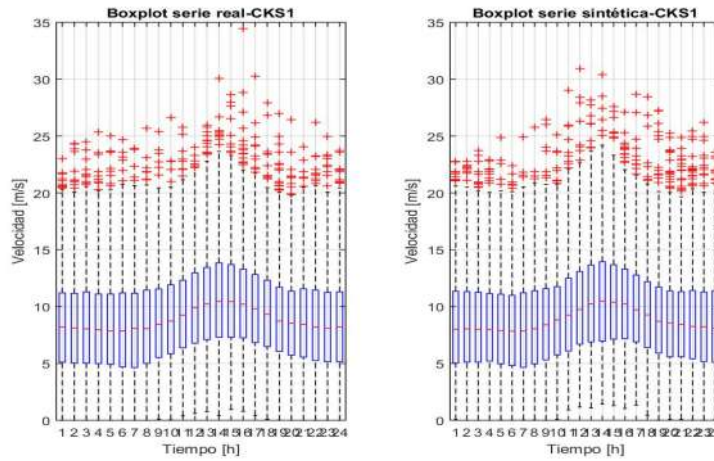


Figura 8.7. Comparación de diagrama de caja para variaciones horarias en sitio CKS1 utilizando modelo ARMA(3,3)

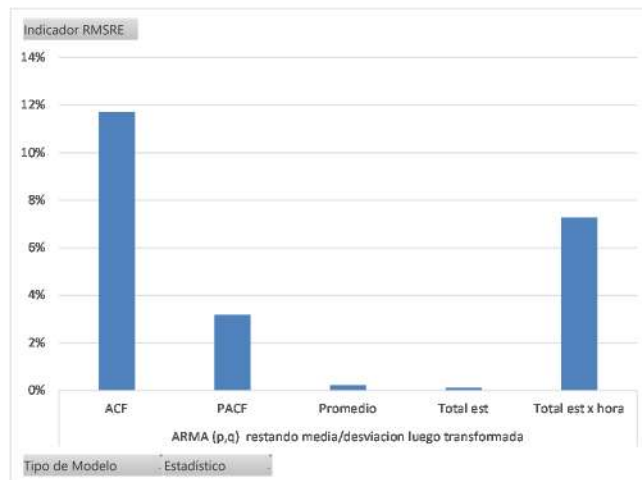


Figura 8.8. Gráfica de indicador para distintos estadísticos en sitio CKS1 utilizando modelo ARMA(3,3)

9 Anexo 3: Modelos espaciotemporales

9.1 Antecedentes generales del anexo

En este anexo comentan en forma resumida los algoritmos para generar series sintéticas que consideran dependencia espacio-temporal. Los modelos seleccionados son cadenas de VAR, CARMA, PCA+ARMA y teoría de cópula gaussiana, debido a las ventajas que se observaron en el estado del arte de este trabajo. Además se muestran resultados de la aplicación de estos algoritmos a los sitios STH1 y CKS1. En cada sitio considera el recurso eólico y la variable que se analiza es la velocidad de viento. El enfoque de este anexo consiste en la aplicación más que en un análisis matemático riguroso, es a través de los resultados que se verifica la bondad de dichos modelos y los problemas que pueden ocurrir en la aplicación.

9.2 VAR [126], [131][138]

Un modelo VAR es una generalización multivariable del modelo AR que fue descrito en el anexo anterior. Es utilizado cuando se desea modelar la dependencia lineal entre varias series temporales, lo que implica que considera dependencia temporal y espacial. Un modelo k-dimensional VAR(p) para $\vec{Z}_t = [Z_{1,t}, Z_{2,t}, \dots, Z_{k,t}]$ queda definido por:

$$\vec{Z}_t = \sum_{i=1}^p A_i \vec{Z}_{t-i} + \vec{\varepsilon}_t \quad (9.1)$$

Donde A_1, \dots, A_p son las matrices con coeficientes del modelo y $\vec{\varepsilon}_t$ es ruido blanco. Un modelo VAR estandarizado puede ser definido como un modelo VAR con $E(Z_{i,t}) = 0$ y $Var(Z_{i,t}) = 1$ para todo i y t lo que implica que es un proceso estacionario. Además, se distribuye en forma gaussiana.

Un modelo VAR es estable si para todas las raíces del polinomio característico se encuentran fuera del círculo unitario [131], matemáticamente queda escrito por:

$$\det(I_k - A_1 w - A_2 w^2 - \dots - A_p w^p) \quad (9.2)$$

Donde I_k es la matriz identidad de orden k. El modelo $VAR_k(p)$ de la ecuación (9.1) puede ser re escrito como un modelo $VAR_{kp}(1)$ utilizando su formulación en espacio de estados como:

$$\vec{Z}_t = A \vec{Z}_{t-1} + \vec{\varepsilon}_t \quad (9.3)$$

La ecuación (9.3) puede re-escribirse en forma matricial:

$$\begin{bmatrix} Z_t \\ Z_{t-1} \\ Z_{t-2} \\ \vdots \\ Z_{t-p+1} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_k & 0 & \dots & 0 & 0 \\ 0 & I_k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_k & 0 \end{bmatrix} \begin{bmatrix} Z_{t-1} \\ Z_{t-2} \\ Z_{t-3} \\ \vdots \\ Z_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (9.4)$$

La autocovarianza de un modelo VAR(p) puede obtenerse a través de:

$$\Gamma_z(h) = J \left(\sum_{i=0}^{\infty} \left(\vec{A} \Sigma_{\varepsilon} \left(\vec{A}^{i+h} \right)^T \right) \right) J^T \quad (9.5)$$

Donde J es una matriz de orden $k \times kp$ $[J, 0, \dots, 0]$ y Σ_{ε} es la matriz de covarianza de $\vec{\varepsilon}$ [131]. Se definió el modelo VAR estandarizado como aquel modelo que tiene $Var(Z_{i,t}) = 1$. En ese caso, las autocorrelaciones del modelo estandarizado $VAR_k(p)$ vienen dadas por la ecuación:

$$R_z(h) = \Gamma_z(h) = \begin{bmatrix} \rho_z(1,1,h) & \rho_z(1,2,h) & \cdots & \rho_z(1,k,h) \\ \rho_z(2,1,h) & \rho_z(2,2,h) & \cdots & \rho_z(2,k,h) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_z(k,1,h) & \rho_z(k,2,h) & \cdots & \rho_z(k,k,h) \end{bmatrix} \quad (9.6)$$

Donde ρ_z es el coeficiente de correlación y h el retardo. Cuando el retardo $h = 0$ en el modelo estandarizado $\rho_z(i, i, 0) = Var(Z_{i,t}) = 1$, entonces los coeficientes de la diagonal de la matriz $R_z(0)$ son unos y también es simétrica, esto es $\rho_z(i, j, 0) = \rho_z(j, i, 0)$ [138].

Los elementos de la diagonal de la matriz $R_z(h)$ corresponden a los coeficientes de autocorrelación de la ACF asociada a cada serie de tiempo. Los elementos fuera de la diagonal corresponden a los coeficientes de autocorrelación cruzada.

La estructura de la dependencia de un modelo VAR estandarizado puede ser definida analíticamente por $R_z(h)$, la cual entrega la ACF, XCF y la cópula gaussiana. Esta última entrega información respecto a la forma de la estructura. De esa forma $Z_{2,t}, Z_{1,t}$ y $Z_{2,t}, Z_{1,t-2}$ de distribuyen de acuerdo a una distribución bivariada normal [138].

Se ha empezado por los antecedentes de un modelo VAR estandarizado porque es la base de las metodologías propuestas en la literatura para generar series sintéticas. En general se tienen k sitios y para cada sitio se dispone de mediciones con n mediciones. Matemáticamente se tiene el vector multivariable con información de los sitios $\vec{X}_t = [X_{1,t}, X_{2,t}, \dots, X_{k,t}]$. En las distintas metodologías para generar series sintéticas se busca transformar todas aquellas series que no se distribuyan con una normal estandarizada. Esto se puede lograr aplicando la transformada integral según se explicó en el anexo 1.

Los distintos procesamientos para transformar y adecuar los datos a una forma estandarizada se muestran en la Figura 9.1. Se pueden mezclar dichos procesos y como en la metodología que se mencionó en el capítulo 3 se deben aplicar las pruebas de hipótesis para verificar que se cumple la condición.

La serie sintética se genera a partir de un proceso de muestreo al proceso $\vec{\varepsilon}_t \sim N(0, \Sigma_\varepsilon)$. Si es ruido blanco, se debe distribuir de acuerdo a una normal y si no lo es se deberá utilizar la cópula gaussiana para muestrear. Además, se debe verificar que no se encuentran autocorrelacionados.

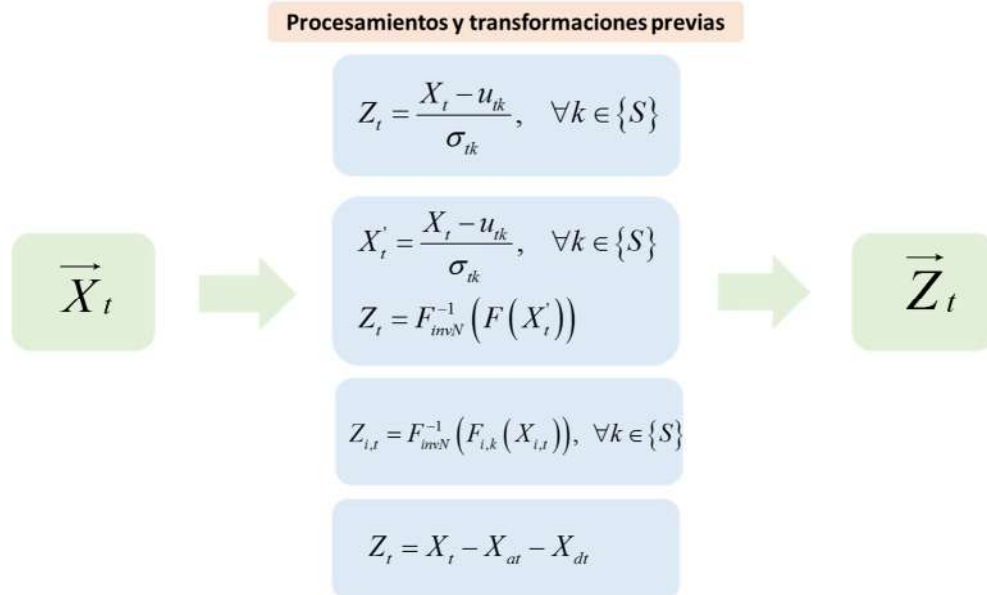


Figura 9.1. Metodologías propuestas para transformar series de tiempo a su forma estandarizada

La estimación del orden se realiza utilizando el criterio AIC y se verifican los supuestos de los residuos. Si estos supuestos no se superan se deberá aumentar el orden p hasta que se pasen los supuestos de autocorrelación. La estimación de los parámetros se realiza utilizando método de los mínimos cuadrados y no se detalla en este trabajo, sólo se aplica el toolbox de [134]. Un mayor detalle respecto a la estimación puede encontrarse en los libros [126], [131].

Existe un aspecto que es muy importante considerar cuando se utiliza esta estructura con series de velocidad de viento y radiación. A la fecha se ha publicado un artículo que considera simultáneamente series de radiación y velocidad de viento [86]. Existe una diferencia cuando se está en presencia de radiación, porque la estimación de los parámetros se aplica al vector \vec{Z}_t y este solo puede ser obtenido a través de la aplicación de la transformada integral, y a su vez, la transformada solo puede ser utilizada cuando la FDA es estrictamente creciente. Esta condición no se da para las horas en las cuales no existe radiación, por lo tanto para tener un vector \vec{Z} que incluya mediciones de velocidades de viento, potencia eólica con radiación para las horas en que no exista radiación se incluye muestra aleatoria no correlacionada. Cuando se genere la serie sintética y se devuelva al espacio real se hace cero todos aquellos valores para los cuales no está presente el recurso solar.

A continuación se detalla el algoritmo cuando se está en presencia del recurso solar dado que existe una diferencia que se agrega el efecto de las horas en las cuales no existe radiación. Respecto a la generación en presencia del recurso eólico se debe seguir el diagrama de flujo de la Figura 3.21.

Algoritmo en presencia de series de recurso eólico y solar

1. Estimar FDA
 - 1.1. Estimar FDA para cada hora para el recurso eólico
 - 1.2. Estimar FDA para todas las horas en que exista el recurso solar
2. Transformar el vector \vec{X}_t aplicando la transformada integral utilizando las FDA del punto anterior
3. Estimar orden p utilizando criterio AIC. El orden p es aquel modelo que tiene el menor AIC
4. Verificar supuestos de autocorrelación de los residuos ¿pasa prueba de hipótesis? Si la respuesta es sí, pasar al siguiente paso, de lo contrario aumentar el orden p en $p+1$ hasta que pase las pruebas de autocorrelación
5. Aplicar pruebas de hipótesis de normalidad y exploración gráfica de datos
6. Si los residuos son normales utilizar muestreo usando factorización Cholesky, de lo contrario utilizar cópula gaussiana
7. Aplicar inversa de FDA
 - 7.1. Aplicar inversa de la FDA para cada hora del recurso eólico
 - 7.2. Aplicar inversa de la FDA para el recurso solar en las horas que exista el recurso. Para las horas en las cuales no exista el recurso solar hacer cero
8. Termina proceso

A continuación se presentan los resultados del modelo escogido aplicando la metodología del caso de estudio. Los resultados que se muestran son aquellos que den claridad a la definición del orden y análisis de residuos. Los estadísticos de la serie sintética no se muestran en esta sección del anexo debido a que se analizan y presentan en el capítulo 4 de este trabajo.

La transformación aplicada a estas series consiste en:

$$Y_t = \frac{X_t - u_{tk}}{\sigma_{tk}}, \quad \forall k \in \{S\} \quad (9.7)$$

$$Z_{i,t} = F_{invN}^{-1} \left(F_{i,k} \left(Y_{i,t} \right) \right), \quad \forall k \in \{S\} \text{ y } \forall i \in \{K\} \quad (9.8)$$

En la ecuación (9.7) u_{tk} y σ_{tk} corresponde a la media y desviación estándar mensual. Una vez aplicada esta ecuación se utiliza la transformada integral por hora. De esta forma se obtiene el vector \vec{Z}_t , el cual es gaussino, estacionario. En la Figura 9.2 se muestra el histograma y el qqplot para el sitio STH1, en estos gráficos se observa que la serie se aproxima a una distribución normal. Aplicando las pruebas de hipótesis χ^2 y Anderson-Darling se ratifica que \vec{Z}_t es gaussina con media cero y varianza uno. Por otro lado, para verificar que la estacional horaria se haya removido se aplica la prueba Kruskas Wallis, los resultados se muestran en la Figura 9.3 y Figura 9.4. El p-valor es mayor a 0.05 y el diagrama de dispersión muestra que para todas las horas la distribución de los datos es idéntica, con ello es posible concluir que las transformaciones permitieron remover las estacionalidades de la serie.

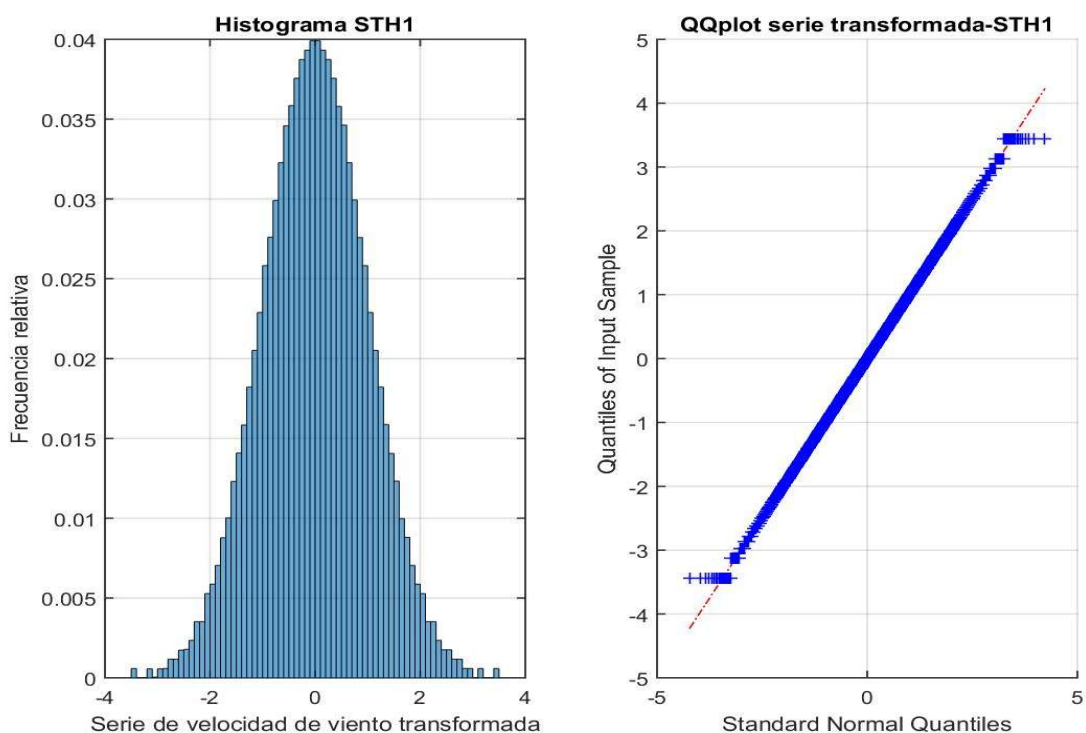


Figura 9.2. Histograma normalizado y qqplot para serie transformada asociada al sitio STH1

Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Groups	225971.1	23	9824.83	0	1
Error	5699820466453.4	40872	139455384.28		
Total	5699820692424.5	40895			

Figura 9.3. Resultado de aplicación prueba de hipótesis Kruskas Wallis a serie transformada asociada a sitio STH1

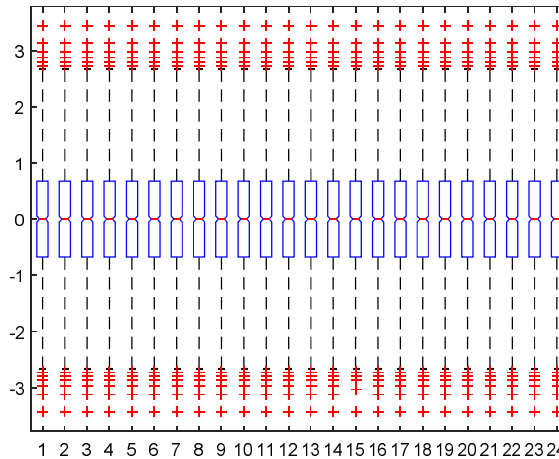


Figura 9.4. Diagrama de cajas sitio STH1 transformado

Con la serie transformada se procede a estimar el orden del modelo utilizando el criterio AIC. El cálculo se realiza utilizando las ecuaciones del anexo 1, sección 7.4. Esta metodología de estimar el orden del modelo es obtenida del libro [126] y la aplicación se hace a través de funciones pre establecidas en el package MTS de R y que está desarrollado en conjunto con el libro. Se escoge el orden de modelo al cual se asocia el menor valor de AIC. Según se muestra en la Tabla 9.1 el orden del modelo sería 16. Sin embargo, este orden es el inicial y se debe verificar que los residuos cumplan con las hipótesis del modelo, o sea que estén autocorrelacionados y sean gaussianos.

Tabla 9.1. Criterio AIC para distintos valores de p

Orden p	AIC
9	-3.96918415
10	-3.96926145
11	-3.96917219
12	-3.96934744
13	-3.96938408
14	-3.96947202
15	-3.96931081
16	-3.96986969
17	-3.96981028
18	-3.96971685
19	-3.96966059
20	-3.9696988

Para verificar si no existe autocorrelación y correlación cruzada se grafican las ACF, PACF y CCF en la Figura 9.5 y Figura 9.6. Para las tres gráficas se observa que los coeficientes de correlación es menor a la banda límite para retardos distintos de cero, con ello se verifica que los residuos cumplen la condición referente a la no autocorrelación para retardos distintos de cero. Además, se muestra en la Figura 9.7 la variación de los p-valores de la prueba de hipótesis Ljung-Box, lo que verifica las conclusiones obtenidas a través de la observación de las ACF, PACF y CCF.

Con el análisis realizado se demuestra que la transformación y orden del modelo son adecuados para las series que se encuentran estudiando.

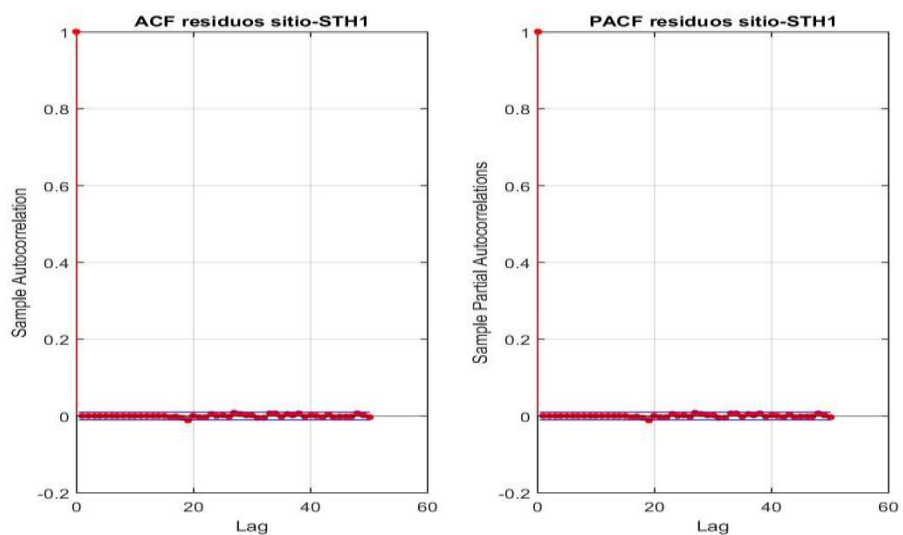


Figura 9.5. ACF y PACF de los residuos asociados al sitio STH1

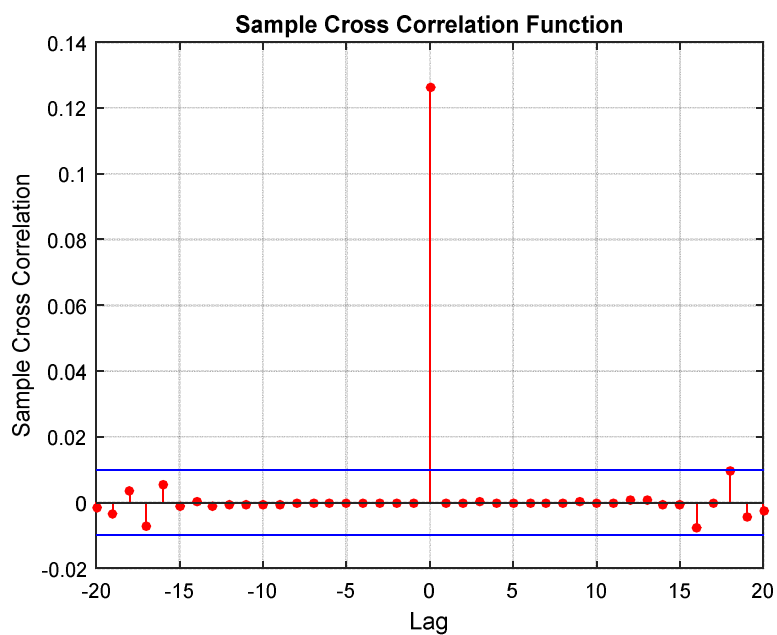


Figura 9.6. CCF de los residuos

p-values of Ljung-Box statistics

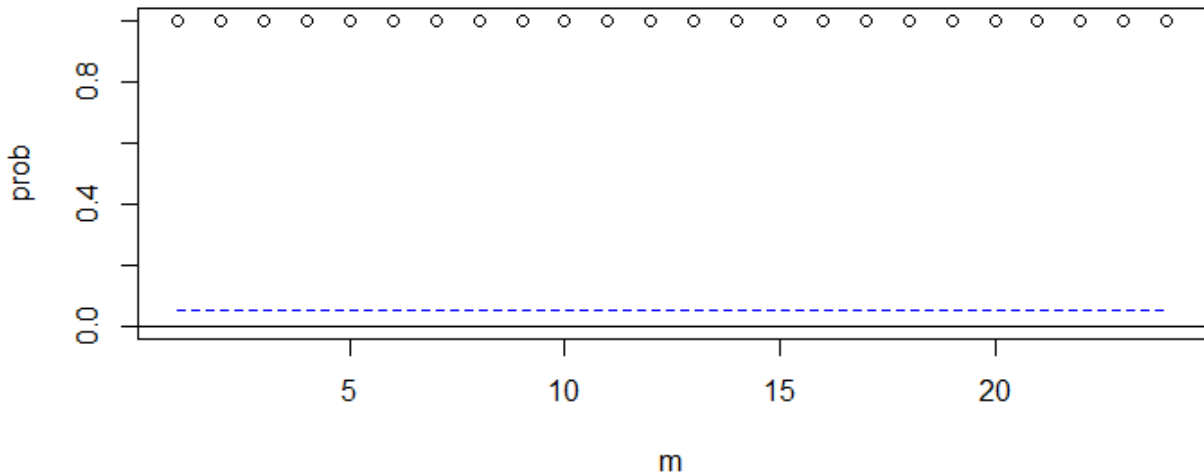


Figura 9.7. Gráfica de los p-valores tras aplicar la prueba de hipótesis Ljung-box

9.3 CARMA [7], [100], [107], [108]

Este modelo multivariable consiste en diagonalizar la matriz de coeficientes A_i de la ecuación (9.9), esto permite que sea posible ajustar un modelo ARMA(p,q) en forma independiente para cada sitio bajo análisis. Una ventaja radica en que la dependencia temporal puede ser caracterizada de mejor manera. La dependencia espacial es introducida a través del vector de ruido blanco, el cual se asume que se distribuye de acuerdo a $N(0, \Sigma)$. Al realizar esta suposición sólo se está modelando la correlación para el retardo cero, lo que sería una desventaja para este modelo.

Las metodologías se basan en los mismos principios que el modelo VAR de la sección anterior. Se tiene un vector \vec{X}_t que no es gaussiano ni estacionario y se debe transformar. Las transformaciones que pueden ser aplicadas son idénticas a las del modelo VAR según se mostró en la Figura 9.1 y en las ecuaciones (9.7) y (9.8).

$$\vec{Z}_t = \sum_{i=1}^p A_i \vec{Z}_{t-i} + \vec{\varepsilon}_t \quad (9.9)$$

Una vez transformada el vector \vec{X}_t se ajusta un modelo ARMA a cada sitio siguiendo la metodología propuesta en el anexo 2. Con los parámetros del modelo se simula la serie sintética y luego se transforma dicha serie al espacio real para tener los datos de velocidad de viento o potencia.

El algoritmo queda de la siguiente forma:

1. Remover estacionalidades
2. Estimar FDA acumulada de la serie sin estacionalidades
3. Transformar \vec{X}_t a \vec{Z}_t utilizando transformada integral
4. Ajustar modelos ARMA para cada sitio
5. Verificar que se cumplan hipótesis sobre los residuos
6. Simular utilizar errores correlacionados a través de la factorización de Cholesky
7. Transformar series sintética a espacio real

Para mostrar la aplicación del algoritmo se muestra su aplicación a las series que fueron estudiadas en el caso de estudio. La estacionalidad de removida utilizando la ecuación (9.10) en donde u_{tk} y σ_{tk}

corresponde medias y desviaciones estándar horarias mensuales, esto es: $24 \times 12 = 288$ valores para la media y la desviación estándar. Luego, a la serie resultante se aplica la transformada integral según se muestra en la ecuación (9.11)

$$Y_t = \frac{X_t - u_{tk}}{\sigma_{tk}}, \quad \forall k \in \{S\} \quad (9.10)$$

$$Z_{i,t} = F_{invN}^{-1} \left(F_i(Y_{i,t}) \right), \quad \forall i \in \{K\} \quad (9.11)$$

En la Figura 9.8 se muestra el histograma de la serie transformada al sitio STH1 así como una curva qqplot. En ambas gráficas se muestra que la transformación permitió que el nuevo proceso sea gaussiano y estacionario. Para cada sitio a través del criterio AIC se estimó el orden de los modelos, quedando:

- Para el sitio STH1 un modelo ARMA(3,3)
- Para el sitio CKS1 un modelo ARMA(3,3)

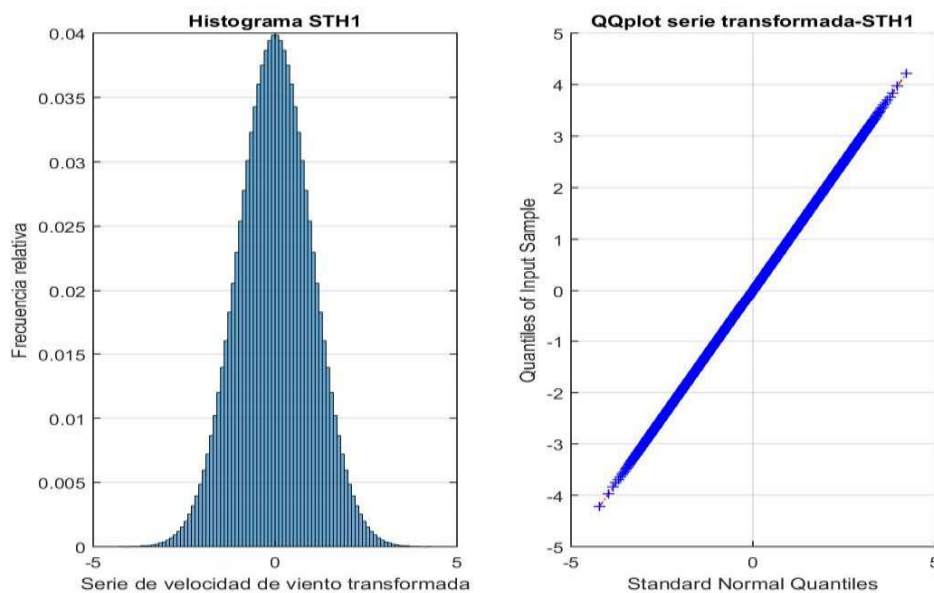


Figura 9.8. Histograma y qqplot para serie transformada del sitio STH1

Al simular y aplicar el proceso inverso se evalúa el modelo en forma gráfica y luego a través de los indicadores. De acuerdo a la Figura 9.10 y Figura 9.11 las funciones ACF y PACF son similares al proceso real, además los histogramas de las series sintéticas muestran pocas diferencias en comparación a la serie real.

Según los indicadores que se muestran en la Figura 9.11 este modelo entrega muy buenos resultados en caracterizar los estadísticos totales y mensuales, en ambos casos el RMSRE no supera el 5%. Donde se aprecia un error significativo es en la función de autocorrelación y función de autocorrelación cruzada llegando esta última a tener un error del 18%. Este resultado es coherente con la suposición que se hace en la diagonalización de la matriz de coeficientes, impidiendo una correcta caracterización para el coeficiente de correlación que se encuentran en retardos distintos a cero y se manifiesta en el error obtenido.

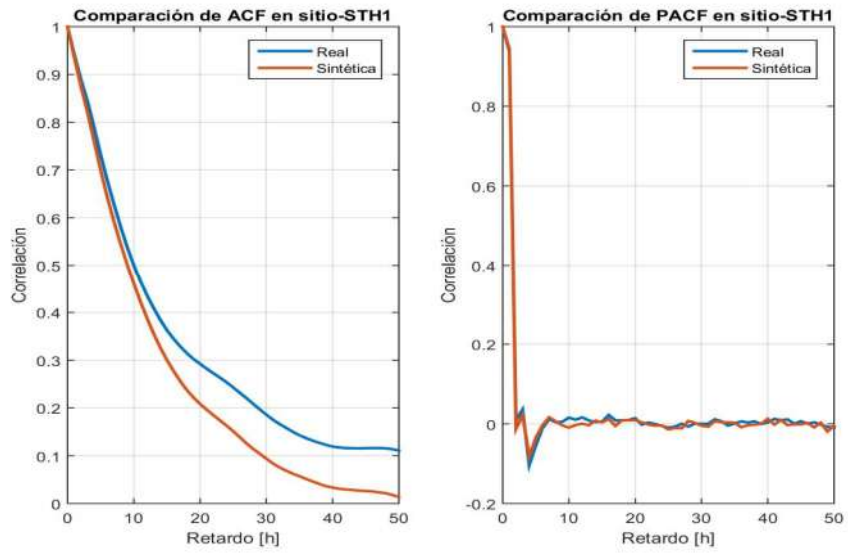


Figura 9.9. Comparación de ACF y PACF entre serie real y sintética para el sitio STH1 utilizando modelo CARMA

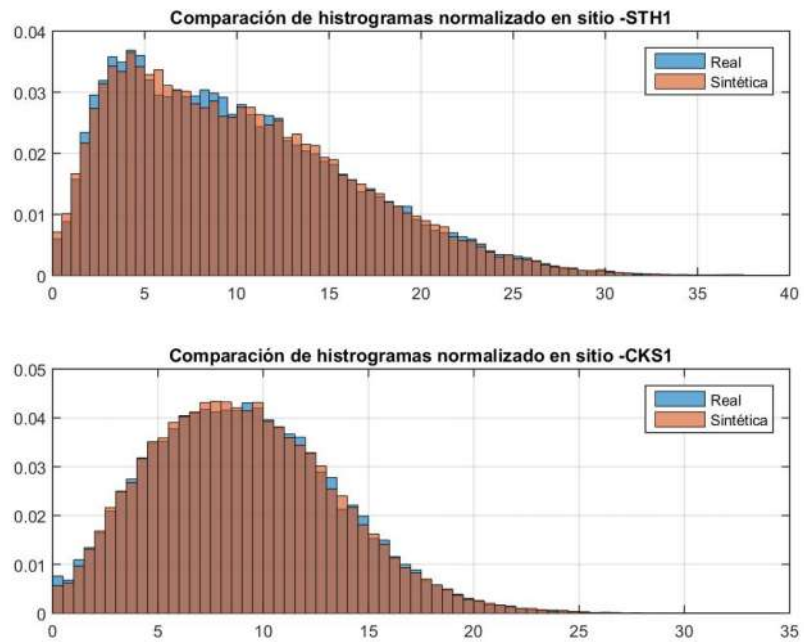


Figura 9.10. Comparación de histogramas entre serie real y sintética para los sitios STH1 y CKS1 utilizando modelo CARMA

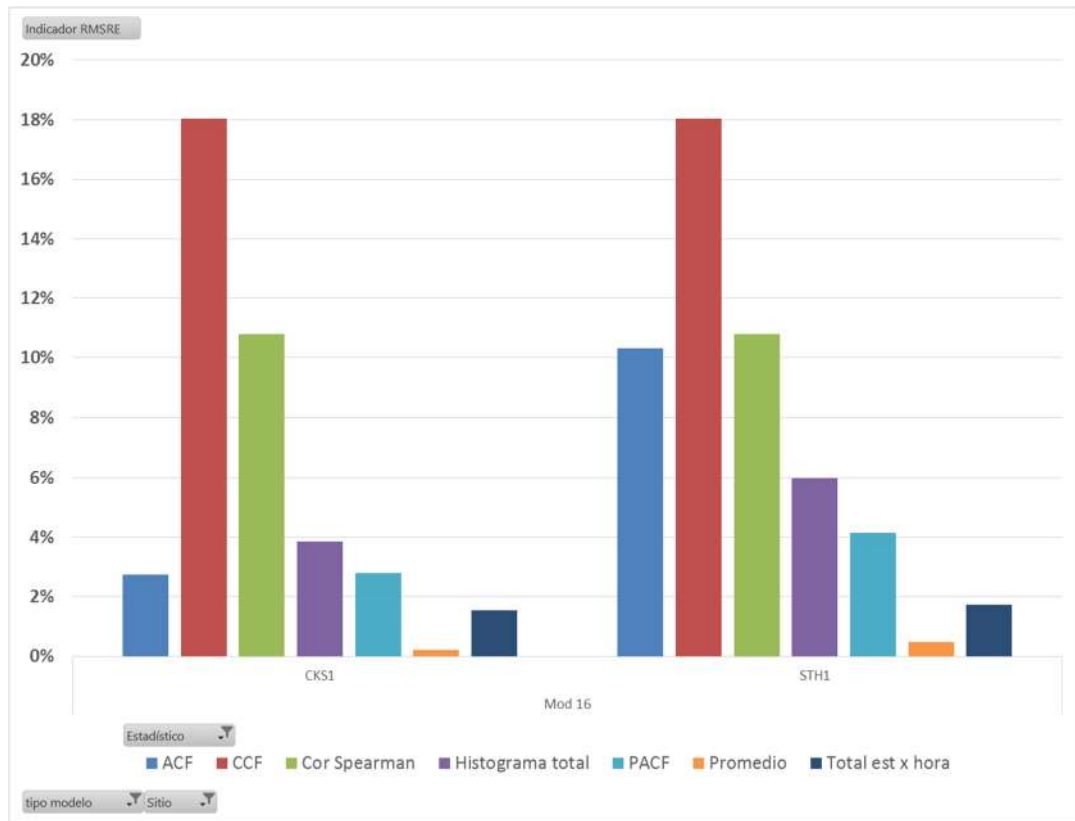


Figura 9.11. Gráfica con indicadores RMSRE utilizando modelo CARMA

9.4 PCA+ARMA [139]

El detalle de este modelo se encuentra en [139], en esta sección se describe e forma general la aplicación de dicha metodología para generar series sintéticas y se presentan resultados para la aplicación del caso de estudio. Este modelo consiste en realizar en transformar una serie que se encuentra correlacionada como lo es el vector \vec{X}_t en un nuevo vector \vec{Z}_t no correlacionado, estacionario y gaussiano. Esta transformación ortogonal se conoce como Análisis de componentes principales, esta técnica se utiliza para el análisis multivariable y es de gran utilidad porque busca extraer información de las series a partir de menos datos, disminuyendo la dimensión del problema.

La aplicación de esta metodología necesita que los datos se encuentren sincronizados como se muestra en la ecuación:

$$\vec{X}_t = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^N \\ x_2^1 & x_2^2 & \cdots & x_2^N \\ \vdots & \vdots & \ddots & \vdots \\ x_s^1 & x_s^2 & \cdots & x_s^N \end{bmatrix} \quad (9.12)$$

Donde S corresponde a la cantidad de la muestra y N la cantidad de sitios, para dicha matriz se define:

$$\mu_s = \frac{1}{N} \sum_{h=1}^N x_s^h \quad (9.13)$$

$$\bar{X}_c = \bar{X}_t - \mu \quad (9.14)$$

La transformación ortogonal que se denomina PCA viene dada por la ecuación:

$$Z = U^T X_c \quad (9.15)$$

Aplicando la ecuación anterior se tiene la serie Z, a la cual se le ajusta un modelo ARMA para cada sitio, recordando que tras realizar la transformación quedan las series no correlacionadas. Una vez aplicada la transformación se puede obtener la serie de recurso a través de la siguiente ecuación

$$X = u + UZ \quad (9.16)$$

La transformación es aplicada en forma directa a través de una función que viene incluida en MATLAB. El algoritmo para generar series sintéticas a través de esta metodología se describe a continuación:

1. Remover estacionalidades utilizando la ecuación (9.10)
2. Utilizar la transformada integral para convertir a serie gaussiana
3. Descomponer utilizando análisis de componentes principales
4. Ajustar orden de modelo ARMA para cada sitio utilizando criterio AIC
5. Verificar que se cumplan hipótesis de autocorrelación sobre los residuos de cada modelo. Si no se cumple aumentar el orden de p hasta que esta condición se cumpla. Al terminar pasar al siguiente paso
6. Simular series sintéticas para longitud arbitraria
7. Volver al plano real aplicando ecuación (9.16) y (9.10)

A continuación se presentan los resultados de aplicar la metodología descrita en esta sección a las series de tiempo asociados a los sitios STH1 y CKS1. Una vez removida la estacionalidad horaria mensual y aplicando la transformada integral se logra un proceso gaussiano con media cero y varianza uno, y no correlacionado. Este hecho se muestra en la Figura 9.12, los histogramas muestran tener la forma de una función de densidad normal y el diagrama de dispersión tiene la forma idéntica para cuando el coeficiente de correlación es cero y es exactamente el valor que toma la correlación para estas series.

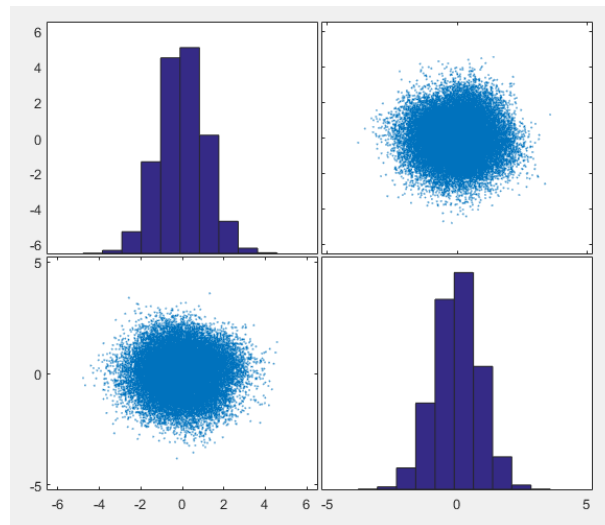


Figura 9.12. Histograma y diagrama de dispersión para series transformadas (eje x sitio CKS1, eje y sitio STH1)

El modelo para cada sitio es ARMA(20,4) para el sitio STH1 y ARMA(16,3) para el sitio CKS1). Las series simuladas muestran que caracterizan adecuadamente las funciones de densidad de probabilidades según se muestra en la Figura 9.13. Por otro lado, la dependencia temporal pareciera quedar bien representada según se muestran las ACF y PACF en la Figura 9.14. En la ACF son idénticas hasta llegar hasta el retardo ocho

aproximadamente, después aparece un error, pero se mantiene la dependencia al tener los valores máximos en los mismos retardos.

Existe un aspecto muy interesante de este modelo. Por un lado, permite ajustar por sitio un modelo ARMA y con ello buscar un orden más complejo para caracterizar de mejor forma estructuras de dependencia temporal, pero la transformación usando PCA tal y como se ha expuesto en el artículo [139] el coeficiente de correlación que desaparece es el correspondiente al retardo cero. Esto se puede apreciar al graficar la función de autocorrelación cruzada de la serie transformada, la cual se muestra en la Figura 9.16, donde el coeficiente de correlación en el retardo cero es nulo, pero si existe correlación significativa para distintos retardo. Al realizar una comparación entre las funciones de autocorrelación cruzada de la serie real y sintética se observa que la serie sintética es capaz de caracterizar adecuadamente la correlación para el retardo cero, pero a medida que aumentan los retardos, aumenta el error.

La aplicación de PCA puede permitir utilizar menos series que la cantidad original y obtener la misma información. Por lo tanto, si bien este modelo falla en modelar adecuadamente toda la estructura de dependencia espacial, puede ser usado cuando se enfrenten situaciones en las cuales la cantidad de sitios sea tan grande que no pueda aplicarse un modelo multivariable.

Los indicadores que se observan en la Figura 9.17 ratifican los comentarios expuestos respecto a las bondades de la utilización de este modelo.

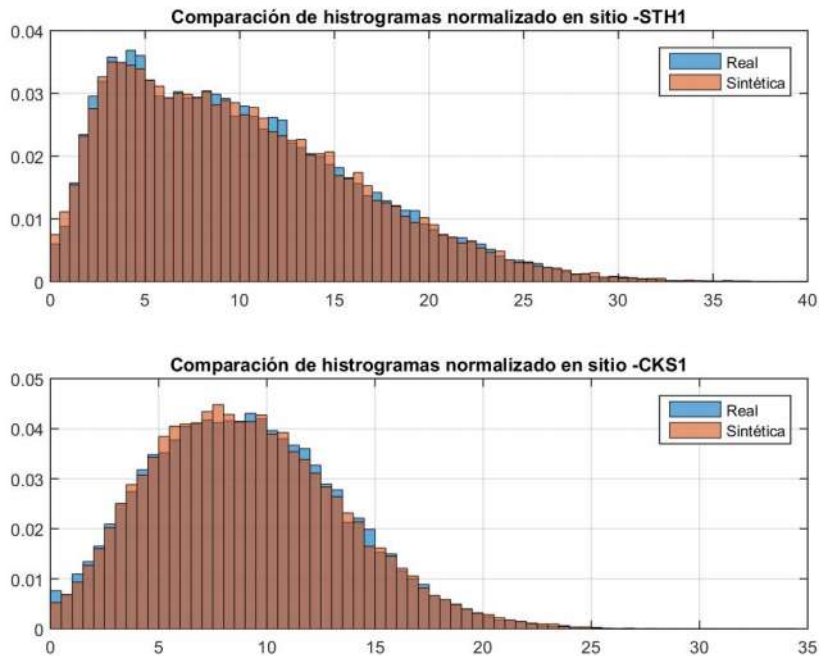


Figura 9.13. Comparación de histogramas entre serie real y sintética para los sitios STH1 y CKS1 utilizando PCA+ARMA

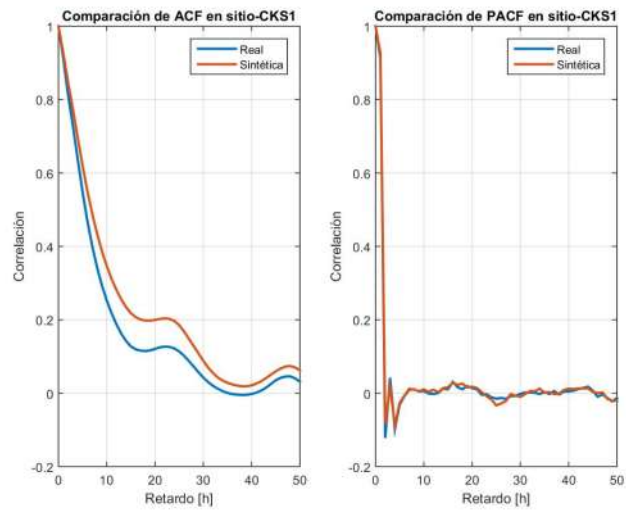


Figura 9.14. Comparación de ACF y PACF para el sitio CKS1 utilizando PCA+ARMA

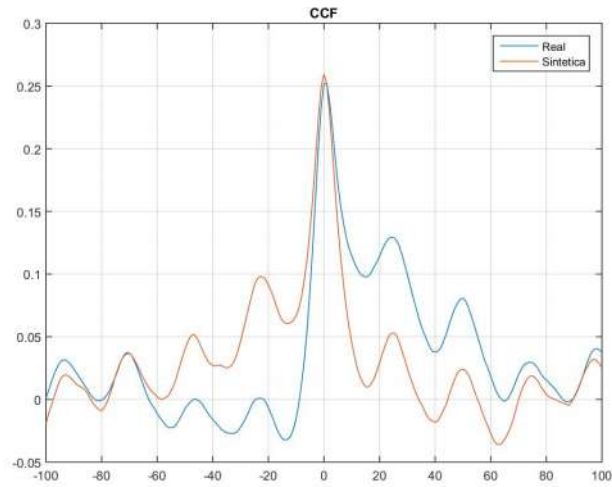


Figura 9.15. Comparación de función de autocorrelación cruzada utilizando PCA+ARMA

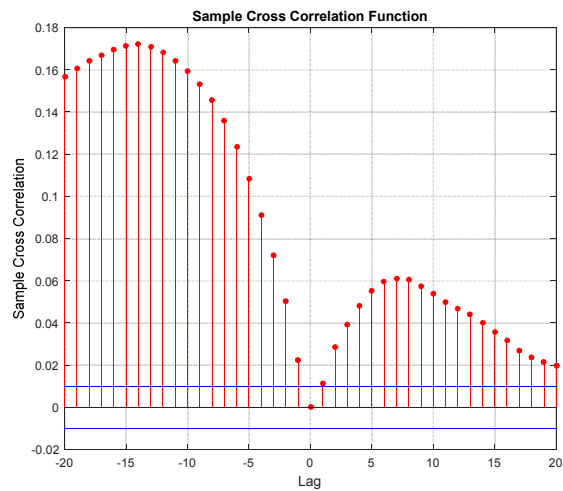


Figura 9.16. Función de autocorrelación cruzada de serie transformada utilizando PCA

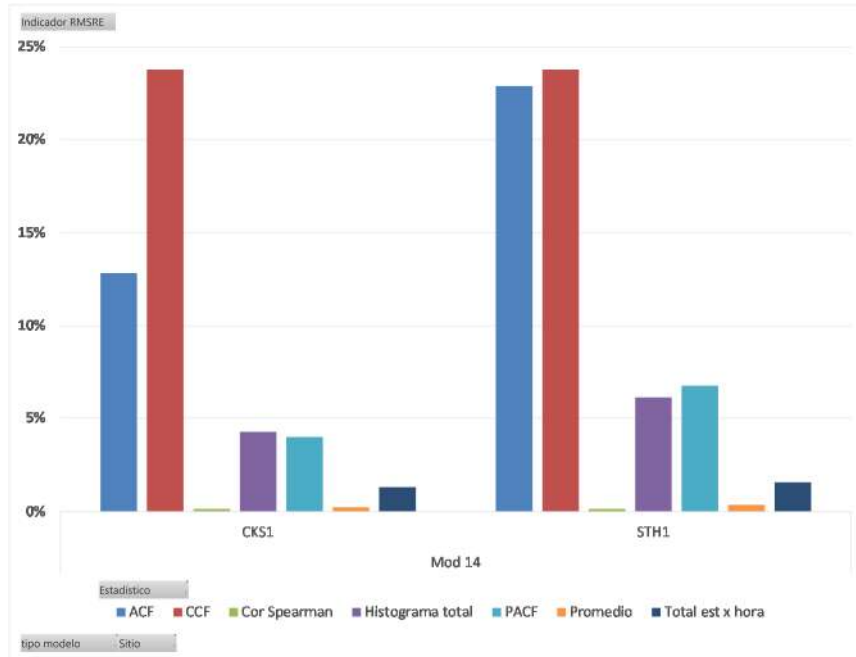


Figura 9.17. Gráfica de RMSRE para distintos estadísticos utilizando PCA+ARMA

9.5 Modelos teoría de Cópula Gaussiana [128]

Una de las cópulas más utilizadas es la cópula Gaussiana. Se define la función de distribución de U como:

$$C_{\Sigma}(u) = H_{\Sigma} \left[F_N^{-1}(u_1), F_N^{-1}(u_2), \dots, F_N^{-1}(u_k) \right] \quad (9.17)$$

Donde F_N^{-1} es la función inversa de la función de probabilidad acumulada (FPA) y H_{Σ} es la función de distribución conjunto (FDC) de la FPA normal de orden k con vector media cero y matriz de covarianza Σ igual a la matriz de correlación.

La función de Cópula gaussiana es definida a través de Σ , la cual puede ser estimada transformando U a:

$$Z = \left[F_N^{-1}(u_1), F_N^{-1}(u_2), \dots, F_N^{-1}(u_k) \right] \quad (9.18)$$

Y luego calcular la matriz de correlación de Z . La transformación presentada en la ecuación (9.18) preserva la correlación de Spearman debido que la inversa de la FPA es una función no decreciente. Por lo tanto, $Z_{ranked} = U_{ranked} = Y_{ranked}$.

Cuando se simula un vector \tilde{U} de la cópula gaussiana, una muestra \tilde{Z} es obtenida de una FDC normal de orden k . Luego la muestra es transformada a:

$$\tilde{U} = \left[F_N(\tilde{Z}_1), F_N(\tilde{Z}_2), \dots, F_N(\tilde{Z}_k) \right] \quad (9.19)$$

Finalmente \tilde{U} es transformada a \tilde{Y} (serie en unidades reales) a través de la transformada integral. La aplicación de esta metodología requiere que la muestra o mediciones dispongan de los datos suficientes para estimar adecuadamente las funciones de distribución para cada variable aleatoria con el propósito de poder aplicar la transformada integral. Además se debe definir las variables aleatorias que serán analizadas. En la práctica se dispone de las series de tiempo $X_{t1}, X_{t2}, \dots, X_{tk}$, siendo k la cantidad de sitios y el conjunto de variables aleatorias que se sea analizar es dependiente de los datos y los objetivos del estudio. Por ejemplo, en la aplicación de esta teoría a la generación de series sintéticas se han utilizado 24 variables para observar variabilidad horaria durante un día, 96 variables para observar estacionalidad, donde se escogen 24 variables

por estación [109][95]. Después que se escogen las variables aleatorias se debe separar las muestras en n conjuntos, siendo n la cantidad de variables aleatorias. La división se realiza para estimar la matriz de correlación entre todas las variables con sus respectivas FDA. Con esta información es posible iniciar el proceso de simulación a través de la cópula gaussiana y luego aplicar la inversa de FDA para volver a tener los datos de viento, radiación o la variable que se esté analizando.

El algoritmo puede ser descrito de la siguiente manera:

1. Definir el conjunto de variables aleatorias que serán analizadas para todos los sitios.
2. Separar la serie de tiempo en n grupos
3. Para cada $i = 1, \dots, n$ estimar FDA, su inversa y matriz de correlación utilizando coeficiente de Spearman (ver Anexo 1)
4. Simular utilizando proceso de muestreo de función de distribución multivariable normal como se mostró en la ecuación (9.19)
5. Aplicar la FDA normal estándar a las variables simuladas
6. Aplicar la inversa de la FDA empírica a los valores simulados del paso anterior para obtener la serie sintética en el espacio real
7. Ordenar los n grupos simulados en orden cronológica

Para mostrar la aplicación del algoritmo para generar series sintéticas utilizando la cópula gaussiana se aplica a las series de tiempo asociados a los sitios STH1 y CKS1. Son estos los resultados que son utilizados en el capítulo 4 comparar la metodología de selección.

Debido a que existe 1 año de datos se definen 24 variables aleatorias por sitio, eso hace un total de 48 variables. Recordando que para las series analizadas la variación estacional horaria es más marcada que las variaciones estacionales mensuales. En la Figura 9.18 se muestra la matriz de correlación entre las 48 variables aleatorias, son 24 para el sitio STH1 y otras 24 para el sitio CKS1.

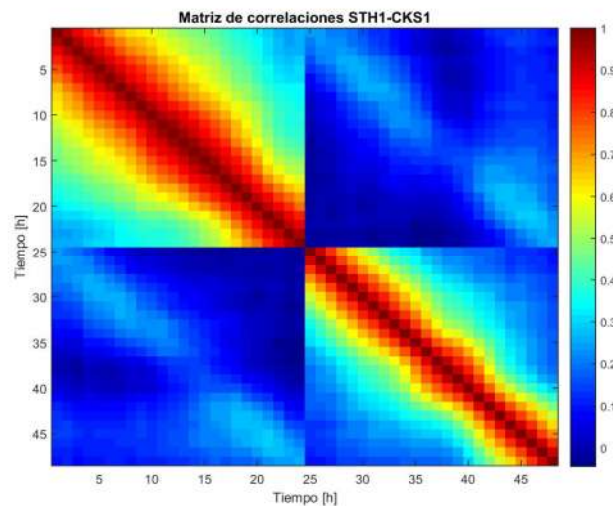


Figura 9.18. Matriz de correlación para 48 variables aleatorias

En la Figura 9.19 se muestran las FDA por cada hora y sitio. A través de estas funciones es posible realizar el paso inverso descrito a través de la ecuación (9.19). La simulación se realiza utilizando MATLAB a través de la función “copularnd”, la cual utiliza la matriz de correlación como dato de entrada.

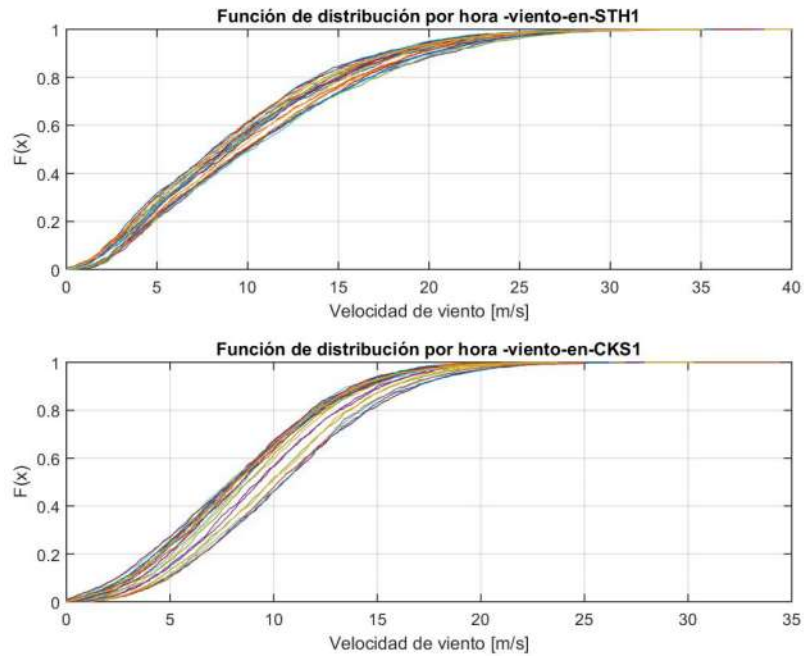


Figura 9.19. Funciones de distribución acumulada para cada sitio y hora

La función de densidad de probabilidades de la serie sintética utilizando este modelo es caracterizada muy bien según se muestra en la Figura 9.20, ambas tienen curva similar. Respecto a las variaciones horarias, estas pueden ser analizadas visualmente a través de diagramas de cajas horarias. En la Figura 9.21 y Figura 9.22 se muestran los diagramas de cajas asociadas a los sitios STH1 y CKS1. En ambos casos, la serie sintética es capaz de caracterizar las variaciones del recurso a lo largo del día, no solo en la media, sino también en la forma de distribución para cada hora del día, es de esperar dicho comportamiento dado que fueron consideradas 24 variables aleatorias por sitio.

El modelo no representa bien la función de autocorrelación, lo que implica que la dependencia temporal no es caracterizada adecuadamente. Esto no significa que la teoría de cópula no es un buen modelo, sino más bien que para las características particulares de la serie, este modelo no representa estas funciones.

En la Figura 9.24 se muestra la matriz de correlación de la serie sintética, al realizar una comparación con la matriz de correlación de la Figura 9.18 es posible concluir que la distribución de las correlaciones son similares debido a la consideración realizada respecto a la cantidad de variables aleatorias. Sin embargo, en un contexto de series de tiempo esta matriz no entrega suficiente información y al utilizar la serie sintética se deberá tener presente que la persistencia temporal no está siendo bien considerada, según fue observado en las ACF y PACF.

Los comentarios y conclusiones respecto a la aplicación de cópula gaussiana a estas series de tiempo es resumida a través del indicador RMSRE. En la Figura 9.25 se muestra una gráfica que muestra la aplicación de este indicador a los siguientes estadísticos: ACF, CCF, coeficiente de correlación de Spearman, histograma, promedio horario y todos los estadísticos horarios. El RMSRE elevado se encuentra para ACF y CCF lo que concuerda con las gráficas de las ACF en ambos sitios, llegando a un 35%. El indicador para los estadísticos horarios tiene un error menor al 5% comprobando que el modelo es capaz de caracterizar los estadísticos horarios adecuadamente.

Una nota debe hacerse en la aplicación de este modelo, el RMSRE aplicado al coeficiente de correlación de Spearman es menor del 5%, lo que se podría interpretar que el modelo está caracterizando adecuadamente la correlación espacial. Sin embargo, si se amplía el análisis a la matriz de autocorrelación se obtiene el

verdadero resultado y se concluye que no es muy buen modelo para estas series de tiempo, en especial en caracterizar la dependencia espacial.

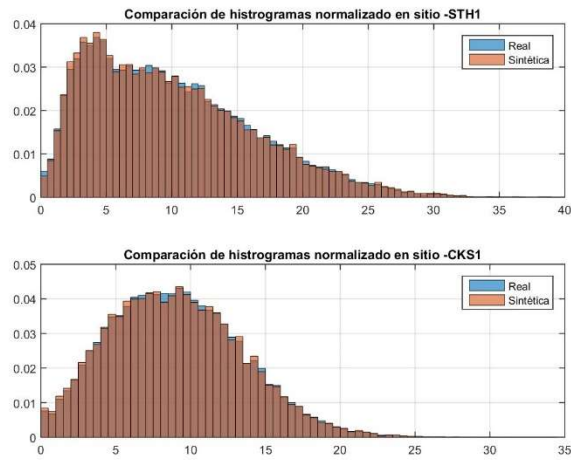


Figura 9.20. Comparación entre serie real y sintética para los sitios STH1 y CKS1 usando una cópula gaussiana

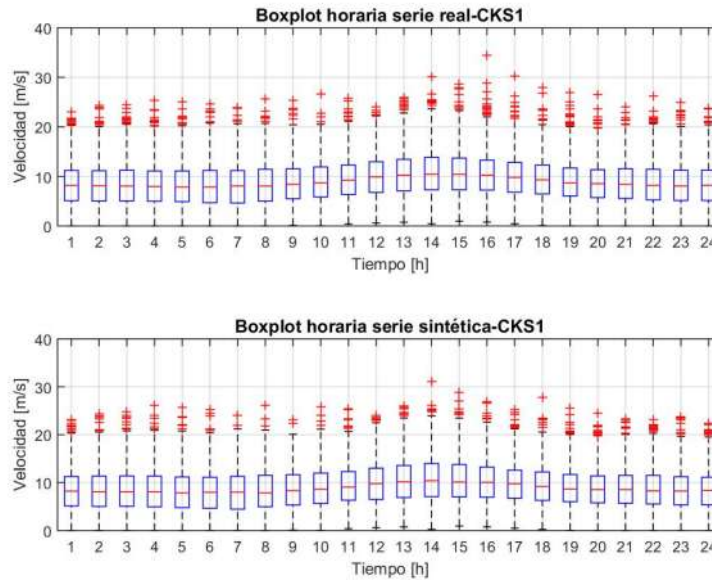


Figura 9.21. Comparación de Boxplot horaria para sitio CKS1 utilizando cópula gaussiana

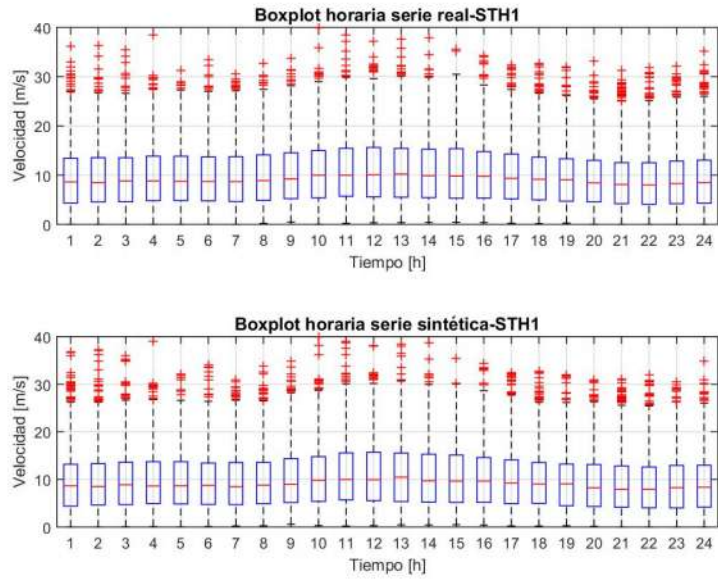


Figura 9.22. Comparación de Boxplot horaria para sitio STH1 utilizando cópula gaussiana

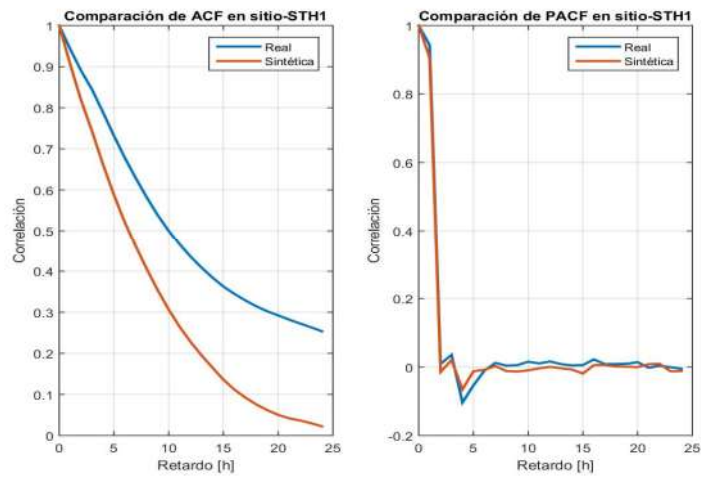


Figura 9.23. Comparación de ACF y PACF para sitio STH1 utilizando cópula gaussiana

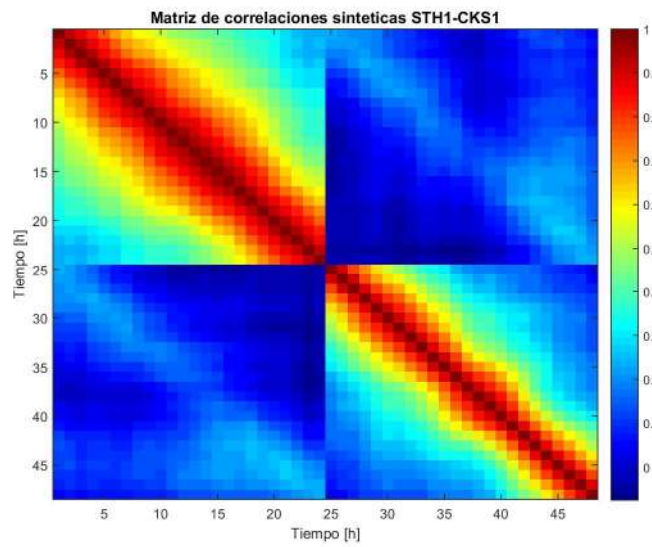


Figura 9.24. Gráfica de matriz de correlación utilizando cópula gaussiana

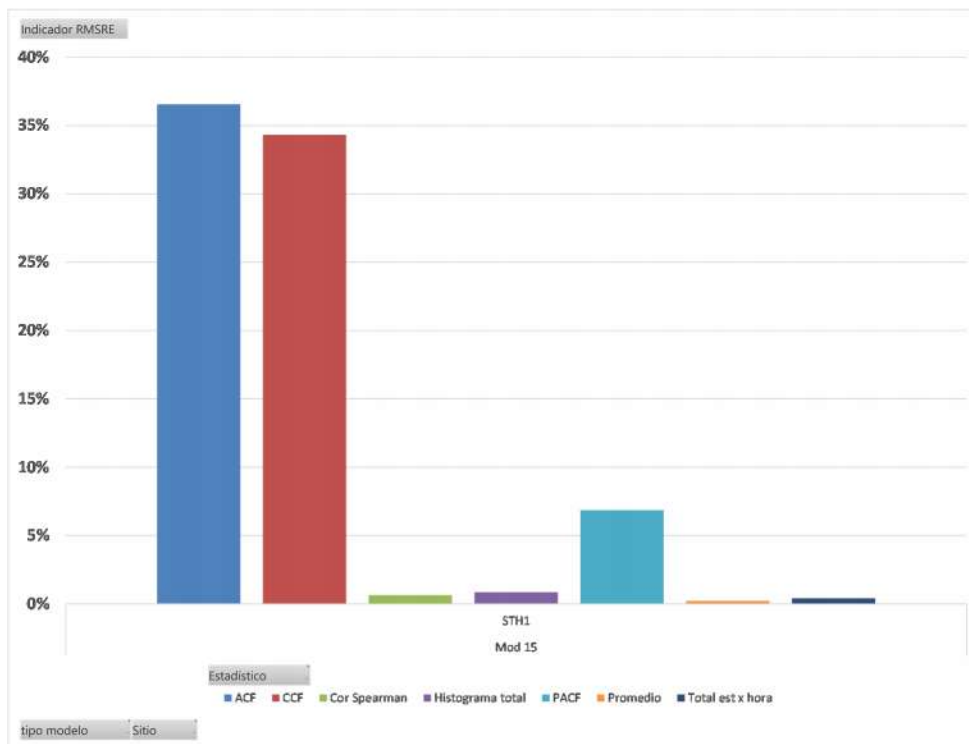


Figura 9.25. RMSRE para sitio STH1 usando cópula gaussiana

10 Anexo 4: Listado de modelos para generar series sintéticas

En esta sección se presenta una serie de tablas que contienen los modelos revisados en el estado del arte. Se presenta cada modelo tipo de dependencia, recurso y resolución considerada en los artículos. Estas tablas se consideran en la metodología propuesta para seleccionar modelos.

Tabla 10.1. Listado de modelos para generar series sintéticas reportados en la literatura parte n° 1

N°	Tipo de dependencia que el modelo considera	Nombre del modelo	Recurso	Resolución temporal	Consideraciones en la aplicación del modelo	Fuente
1	Dependencia temporal	Cadenas de Markov Homogéneas	Velocidad de viento	8 horas	Estacionaria Series no Gaussianas	[77]
2	Dependencia temporal	Cadenas de Markov Homogéneas	Velocidad de viento	1 hora	Estacionaria Series no Gaussianas	[48], [52]
3	Dependencia temporal	Cadenas de Markov Homogéneas	Velocidad de viento	1 hora	Estacionaria Series no Gaussianas	[50]
4	Dependencia temporal	Cadenas de Markov Homogéneas	Potencia eólica de una turbina Velocidad de viento	20, 30 minutos 1 hora	Estacionaria Puede aplicarse a series no Gaussianas	[53],[54]
5	Dependencia temporal	Cadenas de Markov Homogéneas	Potencia eólica de una región	15 minutos	Estacionaria Series no Gaussianas	[54]
6	Dependencia temporal	Cadenas de Markov no Homogéneas	Velocidad de viento	1 hora	Series no Gaussianas Variaciones estacionales mensuales No considera variaciones horarias	[58]

Tabla 10.2. Listado de modelos para generar series sintéticas reportados en la literatura parte n° 2

N°	Tipo de dependencia que el modelo considera	Nombre del modelo	Recurso	Resolución temporal	Consideraciones en la aplicación del modelo	Fuente
7	Dependencia temporal	Cadenas de Markov no Homogéneas	Radiación global horizontal	1 minuto	Serie no gaussiana Variaciones estacionales diarias	[60],[140]
8	Dependencia temporal	Cadenas de Markov	Radiación global horizontal	1 día	Variaciones estacionales Serie no gaussiana	[141]
9	Dependencia temporal	Cadenas markovianas semi	Velocidad de viento	10 minutos	Estacionaria Serie no gaussiana	[112]
10	Dependencia temporal	ARMA (p,q)	Velocidad de viento	1 hora	Estacionaria Serie gaussiana	[142],[74][77]
11	Dependencia temporal	ARMA (p,q)	Velocidad de viento	1 hora	Estacionaria Variación estacional	[113], [34], [82]
12	Dependencia temporal	ARIMA(p,d,q)	Parque eólico	1 hora	Estacionaria Especial aplicación a series de potencia eólica	[68]
13	Dependencia temporal	ARMA(p,q) por mes	Velocidad de viento	1 hora	Estacionaria por mes Variación estacional mensual No es apropiado si existen variaciones horarias	[69]
14	Dependencia temporal	Modelo compuesto	Velocidad de viento	1 hora	Variaciones estacionales Serie no gaussiana	[98],[97]

Tabla 10.3. Listado de modelos para generar series sintéticas reportados en la literatura parte n° 3

N°	Tipo de dependencia que el modelo considera	Nombre del modelo	Recurso	Resolución temporal	Consideraciones en la aplicación del modelo	Fuente
15	Dependencia temporal	ARMA(p)	Potencia eólica de un parque	1 hora	Variación estacional mensual, horaria Serie no gaussiana	[143][46]
16	Dependencia temporal	Copula	Velocidad de viento	10 minutos	Variación horaria y mensual Serie no gaussiana Necesita varios años para poder considerar variaciones mensuales	[95]
17	Dependencia espacio temporal	VAR(p)	Velocidad de viento	1 hora	Serie no Gaussiana Considera variaciones estacionales, asume que al remover estacionalidad, la nueva serie es Gaussiana	[129],[34]
18	Dependencia espacio temporal	VAR(p)	Velocidad de viento	1 hora	Serie no Gaussiana Variación horaria usando transformada integral Considera variaciones horarias mensuales	[100],[7], [105]
19	Dependencia espacio temporal	VAR(p) Transformada integral para modelar variaciones estacionales horarias	Velocidad de viento Radiación solar	1 hora	Serie no Gaussiana Variación horaria usando transformada integral No considera variaciones estacionales mensuales	[114],[115],
20	Dependencia espacio temporal	VAR(p) Transformada integral para modelar variaciones estacionales horarias	Velocidad de viento Radiación solar	1 hora	Serie no Gaussiana Variación horaria usando transformada integral Variaciones estacionales mensuales	[114], más elaboración propia
21	Dependencia espacio temporal	VAR(p)	Potencia eólica de parques	1 hora	Variaciones horarias y mensuales Series no gaussianas Tendencias	[104]
22	Dependencia espacio temporal	CARMA(p,q) con matriz de coeficientes diagonalizada y considera elementos de matriz de correlación	Velocidad de viento	1 hora	Series no gaussianas	[117]

Tabla 10.4. Listado de modelos para generar series sintéticas reportados en la literatura parte n° 4

N°	Tipo de dependencia que el modelo considera	Nombre del modelo	Recurso	Resolución temporal	Consideraciones en la aplicación del modelo	Fuente
23	Dependencia espacio temporal	CARMA(p,q) con matriz de coeficientes diagonalizada y considera elementos de matriz de correlación	Velocidad de viento	1 hora	Serie no gaussianas Variaciones estacionales horarias y mensuales Metodología para generar curva típica de parque eólico	[9],[144]
24	Dependencia espacio temporal	CARMA (p), considera modelo AR(p)	Velocidad de viento	1 hora	Serie no gaussiana Variaciones estacionales horarias y mensuales	[7][101]
25	Dependencia espacio temporal	ARMA (p,q) más teoría de Copula	Velocidad de viento	1 hora	Serie no gaussiana Variaciones estacionales horarias y mensuales	[108],[116]
26	Dependencia espacio temporal	Análisis de componentes principales más ARMA(p,q)	Velocidad de viento	1 hora	Variaciones estacionales horarias y mensuales Serie no gaussiana	[110]
27	Dependencia espacio temporal	Teoría de Copula Vine-Copula	Potencia eólica Velocidad de viento	10 minutos	Serie no gaussiana Variaciones estacionales Necesita de varios años de datos para ajustar el modelo	[11]
28	Dependencia espacio temporal	Teoría de Copula (Diagonal band Copula)	Demanda, velocidad de viento y radiación solar	1 hora	Serie no gaussiana Variaciones estacionales Necesita de varios años de datos para ajustar el modelo	[109]

11 Anexo 5: Pruebas de hipótesis para autocorrelaciones

11.1 Modelo univariable [126]

Una prueba de hipótesis global para que los primeros h coeficientes son cero (h debe ser grande) es el contraste de Ljung-Box. Si la serie se distribuye con una normal, entonces los coeficientes de correlación estimados son asintóticamente normales, con media cero y varianza $(T - k)/T(T + 2)$.

Hipótesis nula $H_0: \rho_1 = \dots = \rho_h = 0$

Hipótesis alternativa $H_a: \rho_i \neq 0$ para algún i que satisface $1 < i \leq h$

El estadístico es:

$$Q(h) = T(T + 2) \sum_{j=1}^h \frac{\hat{\rho}_j^2}{T + j} \quad (9.20)$$

Y se distribuye, asintóticamente, como una χ^2 con grados de libertad igual al número de coeficientes en la suma h menos el número de parámetros estimados, n . Para modelos no estacionales $n=p+q+1$, o $n=p+q$, según que el modelo tenga o no constante, y para los estacionales, que habitualmente no tienen constante, $n=P+p+Q+q$. Se concluye que el modelo es inadecuado si el valor de $Q(h)$ obtenido (9.20) es mayor que el percentil 0.95 de la distribución χ^2 con $h - n$ grados de libertad, que denotaremos por $\chi_{0.95}^2(h - n)$. En general, se rechaza la hipótesis de incorrelación de los residuos cuando la probabilidad:

$$P(\chi^2(h - n) > Q(h)) \quad (9.21)$$

Sea pequeña (menor que 0.05 o 0.001)

11.2 Modelo multivariable [126]

Las hipótesis nula y alternativa son idénticas al caso univariable, con la excepción que en vez de probar los coeficientes de correlación se prueban las matrices de autocorrelación

Hipótesis nula $H_0: \rho_1 = \dots = \rho_h = 0$

Hipótesis alternativa $H_a: \rho_i \neq 0$ para algún i que satisface $1 < i \leq h$

El estadístico Ljung-Box para series multivariadas es:

$$Q_k(m) = T^2 \sum_{l=1}^m \frac{1}{T-l} \text{tr}(\hat{\Gamma}_l \hat{\Gamma}_0^{-1} \hat{\Gamma}_l \hat{\Gamma}_0^{-1}) \quad (9.22)$$

Donde $\text{tr}(A)$ es la traza de la matriz A y T es el tamaño de la muestra. La ecuación anterior puede ser reescrita como:

$$Q_k(m) = T^2 \sum_{l=1}^m \frac{1}{T-l} \hat{b}_l' (\hat{\rho}_l^{-1} \otimes \hat{\rho}_l^{-1}) \hat{b}_l \quad (9.23)$$

Donde $\hat{b}_l = \text{vec}(\hat{\rho}_l')$ y \otimes es el producto Kronecker de dos matrices. Aquí, $\text{vec}(A)$ denota la vectorización columna de la matriz A . Bajo la hipótesis nula $\Gamma_l = 0$ para $l > 0$ y la condición que el vector multivariable x_t se distribuye con una normal, $Q_k(m)$ se distribuye de acuerdo $\chi_{mk^2}^2$ con mk^2 grados de libertad. Se rechaza la hipótesis nula si los valores p de $Q_k(m)$ son menores a un nivel de confianza $1 - \alpha$, con α 0,95 o 0,99.