



UNIVERSIDAD MICHOACANA DE SAN NICOLÁS DE HIDALGO

DIVISIÓN DE ESTUDIOS DE POSGRADO DE LA FACULTAD DE INGENIERÍA MECÁNICA

## MODELO INTEGRAL PARA LA PREDICCIÓN DE LA POTENCIA GENERADA POR EQUIPOS FOTOVOLTAICOS DE GRAN ESCALA

TESIS QUE PARA OBTENER EL GRADO DE DOCTOR EN CIENCIAS  
EN INGENIERÍA MECÁNICA

PRESENTA

**EDUARDO RANGEL HERAS**

**Asesor:** Dr. Erasmo Cadenas Caderón  
**Coasesor:** Dr. Cesar Ángeles Camacho



Morelia, Mich. noviembre de 2018

## Índice de contenido

Índice de contenido.....	2
Índice de tablas .....	7
Índice de figuras .....	8
Resumen .....	10
Abstract .....	11
Objetivo General .....	12
Objetivos específicos .....	13
Justificación .....	14
Hipótesis .....	15
Capítulo I .....	16
Estado del arte .....	16
Capítulo 2 .....	21
Energía Solar .....	21
2.1 Dentro del universo .....	21
2.2. El Sol.....	22
2.2.1 Naturaleza de la radiación solar .....	23
2.3 Geometría del movimiento terrestre .....	24
2.3.1 Ángulos Tierra – Sol .....	26
2.4 Constante solar .....	29
2.5 Variación de la Radiación extraterrestre.....	30
2.6 Radiación fuera de la atmósfera terrestre.....	33
2.6.1 Influencia de la atmósfera terrestre en la radiación solar.....	34
2.6.2 Cálculo de componentes de radiación solar.....	35
2.7 Irradiación solar bajo condiciones de cielo despejado.....	37
Capítulo 3 .....	39
La celda solar.....	39
3.1 El efecto fotovoltaico .....	39
3.1.1 Metales, semiconductores, aislantes .....	40
3.1.2 Dentro de la estructura de banda de un semiconductor.....	42
3.1.3 Absorción de la luz .....	43

<b>3.1.4 Estados permitidos para huecos y electrones</b> .....	45
<b>3.1.5 Distribución de energía para huecos y electrones</b> .....	45
<b>3.1.6 Dopaje</b> .....	47
<b>3.1.7 Transporte del portador</b> .....	49
3.1.7.1 Corriente de deriva.....	50
3.1.7.2 Corriente de difusión.....	52
3.1.7.3 Resistividad del semiconductor.....	54
<b>3.1.8 Ecuaciones fundamentales del semiconductor</b> .....	54
3.1.8.1 Ecuación de Poisson.....	54
3.1.8.2 Ecuación de la continuidad .....	55
<b>3.1.9 Ecuaciones de difusión de portadores minoritarios</b> .....	56
<b>3.1.10 Unión P-N</b> .....	58
<b>3.1.11 Capacitancia de la unión P-N</b> .....	62
<b>3.1.12 La celda Fotovoltaica</b> .....	66
3.1.12.1 Densidad de corriente de los portadores minoritarios.....	66
3.1.12.2 Tasa de generación óptica .....	68
<b>3.1.13 Tasa de recombinación</b> .....	68
<b>3.1.14 Modelo físico de la celda FV</b> .....	69
<b>3.1.15 Tipos de semiconductores</b> .....	70
3.1.15.1 Silicio cristalino.....	70
3.1.15.2 Multicristalino .....	71
3.1.15.3 Amorfo .....	72
3.1.15.4 Película delgada.....	73
3.1.15.5 Célula solar polimérica.....	73
<b>Capítulo 4</b> .....	75
<b>Estadística para datos experimentales</b> .....	75
<b>4.1 Diseño de experimentos.</b> .....	75
<b>4.2. Descripción, exploración y comparación de datos.</b> .....	77
4.2.1. Características importantes de los datos.....	77
4.2.2. Distribuciones de frecuencias.....	78
4.2.3. Visualización de los datos. ....	79
4.2.4. Medidas de tendencia central .....	80

4.2.5. Medidas de variación .....	83
<b>4.3. Tercer y cuarto momento.</b> .....	<b>85</b>
4.3.1. Tercer momento (sesgo).....	85
4.3.2. Cuarto momento (curtosis).....	86
<b>Capítulo 5</b> .....	<b>88</b>
<b>Técnicas econométricas</b> .....	<b>88</b>
<b>5.1. Procesos estocásticos</b> .....	<b>88</b>
5.1.1. Procesos estocásticos estacionarios.....	88
5.1.2. Proceso estocástico de raíz unitaria.....	91
5.1.3. Procesos estocásticos integrados.....	93
<b>5.2. Significancia estadística de los coeficientes de autocorrelación</b> .....	<b>94</b>
<b>5.3. Prueba de raíz unitaria</b> .....	<b>95</b>
5.3.1. La prueba Dickey-Fuller aumentada (DFA) .....	97
<b>5.4. Modelos autorregresivos y de rezagos distribuidos.</b> .....	<b>98</b>
<b>5.5. Prueba de cointegración de Eagle-Granger</b> .....	<b>99</b>
5.5.1. Prueba de Granger.....	99
<b>Capítulo 6</b> .....	<b>102</b>
<b>Técnicas de predicción y análisis de series de tiempo</b> .....	<b>102</b>
<b>6.1. Modelos estadísticos para la predicción de series de tiempo.</b> .....	<b>102</b>
6.1.1. Herramientas básicas de pronóstico. ....	102
6.1.2. Resúmenes gráficos.....	102
6.1.3. Resúmenes numéricos. ....	103
6.1.4. Autocorrelación.....	106
6.1.5. Medición de la precisión de los pronósticos. ....	107
<b>6.2 La metodología Box – Jenkins para modelos ARIMA.</b> .....	<b>108</b>
6.2.1. Examen de correlaciones en los datos de series de tiempo .....	108
6.2.2. Modelo de ruido blanco.....	110
6.2.3. Identificación de la estacionalidad en las series de tiempo. ....	113
6.2.4. Remover la no estacionariedad en una serie de tiempo .....	117
6.2.5. Diferenciación estacional .....	118
6.2.6. Notación de desplazamiento hacia atrás “Backshift” .....	118
6.2.7. Modelos ARIMA para datos de series de tiempo .....	119

<b>6.3 Pruebas de desempeño</b> .....	126
6.3.1 Error medio (ME).....	127
6.3.2 Error medio absoluto (MAE) .....	127
6.3.3 Error cuadrado medio (MSE).....	127
6.3.4 Raíz del error cuadrado medio (RMSE).....	127
<b>Capítulo 7</b> .....	128
<b>Redes neuronales artificiales</b> .....	128
<b>7.1 Una introducción a la computación neuronal</b> .....	128
<b>7.2 Definición de una red neuronal</b> .....	128
<b>7.3 Ventajas de las redes neuronales</b> .....	130
<b>7.4 Fundamentos de las redes neuronales (El modelo biológico)</b> .....	130
7.4.1 Estructura de la neurona.....	130
<b>7.5 Naturaleza bioeléctrica de la neurona</b> .....	131
<b>7.6 Red multicapa con alimentación directa (MLFF)</b> .....	132
<b>7.7 Red no lineal autorregresiva con entradas exógenas (NARX)</b> .....	133
<b>Capítulo 8</b> .....	136
<b>Modelado de sistemas fotovoltaicos</b> .....	136
<b>8.1 Modelado de Sistemas FV</b> .....	136
8.1.1 Modelo de la celda fotovoltaica (FV). .....	136
<b>8.2 Modelado del generador fotovoltaico</b> .....	138
8.2.1 Modelo dependiente de la temperatura .....	140
<b>8.3 Resultado de la potencia generada utilizando el software</b> .....	141
<b>Capítulo 9</b> .....	143
<b>Desarrollo de modelos de pronóstico</b> .....	143
<b>9.1 Análisis estadístico de la serie de tiempo de radiación solar</b> .....	143
<b>9.2. Función de autocorrelación y autocorrelación parcial</b> .....	144
<b>9.3 Modelo estacional ARIMA</b> .....	145
9.3.1 Análisis y monitoreo de los errores modelo ARIMA .....	145
9.3.2 Uso del modelo ARIMA para la predicción 24 horas adelante.....	147
<b>9.4 Modelo de NAR</b> .....	150
9.4.1 Normalización de los datos de entrada.....	150
9.4.2 Análisis y monitoreo de los errores modelo NAR .....	151

9.4.3 Implementación del modelo NAR para pronóstico de 24 horas adelante del año 2011 .....	152
<b>9.5 Modelo NARX .....</b>	<b>155</b>
9.5.1 Determinación de las entradas del modelo NARX .....	155
9.5.2 Normalización de los datos .....	156
9.5.3 Implementación del modelo NARX para pronóstico de 24 horas adelante del año 2011 .....	157
<b>Conclusiones. ....</b>	<b>161</b>
<b>Referencias Biográficas.....</b>	<b>162</b>

## Índice de tablas

<b>Tabla 2. 1.</b> Días promedio recomendados de los meses para promedios de radiación del mes .....	31
<b>Tabla 4. 1.</b> Tabla de frecuencias de la irradiancia solar. ....	78
<b>Tabla 4. 2.</b> Medidas de tendencia central para datos de radiación solar durante las horas del día .....	82
<b>Tabla 4. 3.</b> Medidas de variación para muestras del mes de enero y febrero .....	85
<b>Tabla 4. 4.</b> Coeficiente de asimetría para el mes de enero y febrero datos muestrales de radiación .....	86
<b>Tabla 4. 5.</b> Calculo de la curtosis para las muestras de radiación meses de enero y febrero .....	86
<b>Tabla 6. 1.</b> Radiación solar correspondiente a enero y febrero .....	104
<b>Tabla 6. 2.</b> Cálculo de la auto-covarianza y la auto-correlación utilizando las ecuaciones 6.3 y 6.4, y un rezago de un periodo. ....	106
<b>Tabla 6. 3.</b> .....	107
<b>Tabla 6. 4.</b> Serie de tiempo con 36 valores.....	109
<b>Tabla 6. 5.</b> Patrones esperados en la FAC y la FACP para modelos AR y MA individuales.....	123
<b>Tabla 9. 1.</b> Medidas estadísticas para la serie de tiempo de Temixco, 2010 .....	143
<b>Tabla 9. 2.</b> Medidas estadísticas para la serie de tiempo de Temixco, 2011 .....	144
<b>Tabla 9. 3.</b> Resultados de las pruebas de desempeño, Curva de ajuste .....	146
<b>Tabla 9. 4.</b> Resultado ARIMA de las pruebas de desempeño, Curva de ajuste, Datos predichos, Un día de cada mes .....	147
<b>Tabla 9. 5.</b> Pruebas de desempeño individuales para un día de cada mes .....	148
<b>Tabla 9. 6.</b> Resultados ARIMA de las pruebas de desempeño para cada mes de pronóstico .....	150
<b>Tabla 9. 7.</b> Pruebas de modelo NAR utilizando diferente número de neuronas .....	151
<b>Tabla 9. 8.</b> Resultado NAR de las pruebas de desempeño, Curva de ajuste, Datos predichos, Un día de cada mes. ....	152
<b>Tabla 9. 9.</b> Pruebas de desempeño individuales para un día de cada mes .....	153
<b>Tabla 9. 10.</b> Resultados NAR de las pruebas de desempeño para cada mes de pronóstico .....	153
<b>Tabla 9. 11.</b> Variables meteorológicas Temixco .....	155
<b>Tabla 9. 12.</b> Pruebas del modelo NARX utilizando diferente número de neuronas .....	156
<b>Tabla 9. 13.</b> Pruebas de desempeño individuales para un día de cada mes .....	158
<b>Tabla 9. 14.</b> Resultados NARX de las pruebas de desempeño para cada mes de pronóstico .....	158

## Índice de figuras

<b>Figura 1. 1.</b> Relación entre los horizontes de tiempo, los modelos de predicción y las actividades conexas.....	19
<b>Figura 1. 2.</b> Clasificación de los modelos de predicción basados en resolución espacial de los datos de entrada y resolución temporal.....	20
<b>Figura 2. 1.</b> Trayectoria Sol-Tierra.....	22
<b>Figura 2. 2.</b> La estructura del Sol [43] .....	23
<b>Figura 2. 3.</b> Ángulo de declinación .....	26
<b>Figura 2. 4.</b> Vista de los diferentes ángulos Sol - Tierra .....	27
<b>Figura 2. 5.</b> Variación de las horas de luz solar (HL) con el n-esimo día del año para diferentes latitudes en el hemisferio norte. ....	29
<b>Figura 2. 6.</b> Radiación Sol - Tierra.....	30
<b>Figura 2. 7.</b> Variación de la radiación solar extraterrestre con el día del año .....	31
<b>Figura 2. 8.</b> Dirección de los rayos del Sol a través de la atmósfera.....	32
<b>Figura 2. 9.</b> Variación de la masa de aire con la hora del día.....	33
<b>Figura 3. 1.</b> Configuración experimental para la evaluación del efecto fotoeléctrico .....	40
<b>Figura 3. 2.</b> Representación esquemática de la estructura de banda para: a aislante, b semiconductor, c conductor..	41
<b>Figura 3. 3.</b> Energía versus representación del momentum de la estructura de la banda de energía para un semiconductor de banda prohibida directa.....	43
<b>Figura 3. 4.</b> Energía versus la representación de momentum de la estructura de banda de energía para el semiconductor de banda prohibida indirecto.....	43
<b>Figura 3. 5.</b> Nivel de Fermi y densidad de estado .....	46
<b>Figura 3. 6.</b> Tabla periódica .....	46
<b>Figura 3. 7. a</b> Un dopaje de tipo n con Fosforo As, b dopaje tipo - p con Boro.....	48
<b>Figura 3.8. a</b> Movimiento térmico aleatorio, b movimiento térmico bajo campo eléctrico .....	50
<b>Figura 3. 9.</b> Movimiento de niveles de energía y electrones (agujeros) en presencia de un campo eléctrico .....	51
<b>Figura 3. 10.</b> Semiconductor con campo eléctrico aplicado: se produce una corriente de deriva formada por agujeros y electrones .....	52
<b>Figura 3. 11.</b> Semiconductor con campo eléctrico aplicado: aplicación de la ecuación de continuidad a un volumen con longitud $dx$ .....	55
<b>Figura 3. 12.</b> Representación de una barra semiconductor con una luz incidente en la superficie lateral que provoca la inyección de carga: la concentración de carga se relaja con la constante espacial $L_p$ , para $x$ lo suficientemente lejos de la zona de inyección se restablece la concentración de equilibrio.....	57
<b>Figura 3. 13.</b> Representación esquemática de una unión p-n.....	58
<b>Figura 3. 14.</b> Símbolo eléctrico y voltaje en función del diodo de corriente.....	61
<b>Figura 3. 15.</b> Modelo de circuito físico de una célula fotovoltaica .....	69
<b>Figura 3. 16.</b> Estructura de silicio cristalino (a), policristalino (b) y amorfo (c).....	71
<b>Figura 3. 17.</b> Transferencia de electrones de material donante a receptor (a), sección de una célula orgánica (b).....	72
<b>Figura 4. 1.</b> Sesgo.....	83
<b>Figura 4. 3.</b> Histograma mes de febrero .....	87
<b>Figura 4. 2.</b> Histograma del mes de enero.....	87
<b>Figura 5. 1.</b> Tendencia determinista frente a tendencia estocástica .....	93
<b>Figura 6. 1.</b> Representación esquemática de la metodología Box-Jenkins para modelado de series de tiempo.....	109
<b>Figura 6. 2.</b> Gráfica de tiempo de los datos de la Tabla 6.4.....	110
<b>Figura 6. 3.</b> Gráfica de la Función de autocorrelación para serie de tiempo de la Tabla 6.4 .....	111
<b>Figura 6. 4.</b> Autocorrelaciones parciales para la serie de la Tabla 5.4.....	114

<b>Figura 6. 5.</b> Datos de series de tiempo; (a) serie de tiempo estacionaria en la media; (b) serie de tiempo no estacionaria en la media; y (c) serie de tiempo no estacionaria en la media ni en la varianza .....	115
<b>Figura 6. 6.</b> Índice Dow Jones para 251 días de mercado finalizando el 26 de agosto de 1994.....	117
Figura 7. 1. Esquema de una red neuronal biológica .....	131
<b>Figura 7. 2.</b> RNA, estructura utilizada para obtener la estimación de la radiación solar horaria .....	133
<b>Figura 7. 3.</b> (a) Red neuronal no lineal autorregresiva con entradas externas, (b) Perceptrón Múltiple con una capa oculta [74].....	135
<b>Figura 8. 1.</b> Modelo simplificado de la celda solar .....	137
<b>Figura 8. 2.</b> Modelo de la celda solar con resistencia serie .....	138
<b>Figura 8. 3.</b> (a) Ventana principal y (b) gráficas de curvas de potencia. ....	141
<b>Figura 8. 4.</b> Gráfica temperatura y Radiación solar (b) panel de opciones .....	142
Figura 9. 1 (a) Función de autocorrelación serie de tiempo original, (b) Función de autocorrelación parcial serie de tiempo original, (c) Función de autocorrelación diferenciada 24 rezagos.....	145
<b>Figura 9. 2.</b> (a) Función de autocorrelación errores de la curva de ajuste (b) función de autocorrelación parcial para los mismos errores .....	146
<b>Figura 9. 3.</b> Pronóstico modelo ARIMA de 1 día de cada mes del año 2011, elegido de forma aleatoria .....	147
<b>Figura 9. 4.</b> Columna (a) Resultados del pronóstico, Columna (b) caja de control. Modelo ARIMA .....	149
<b>Figura 9. 5.</b> Función de autocorrelación y autocorrelación parcial de los errores, Modelo NAR .....	151
<b>Figura 9. 6.</b> Pronóstico modelo NAR para 1 día de cada mes del año 2011 elegido de forma aleatoria.....	152
<b>Figura 9. 7.</b> Columna (a) Resultados del pronóstico, Columna (b) caja de control. Modelo NAR.....	155
<b>Figura 9. 8.</b> Función de autocorrelación y autocorrelación parcial de los errores, Modelo NARX .....	157
<b>Figura 9. 9.</b> Pronóstico modelo NARX para 1 día de cada mes del año 2011 elegido de forma aleatoria.....	158
<b>Figura 9. 10.</b> Columna (a) Resultados del pronóstico, Columna (b) caja de control. Modelo NARX.....	160

## Resumen

En el presente trabajo de investigación se establecen las bases para la integración de un modelo de pronóstico de la radiación solar con un modelo para calcular la potencia generada en los sistemas fotovoltaicos. En el caso de los modelos de predicción se utilizaron tres modelos distintos, el modelo estacionario autorregresivo integrado de medias móviles (sARIMA), el modelo no lineal autorregresivo de redes neuronales (NAR) y el modelo no lineal autorregresivo con entradas exógenas de redes neuronales (NARX). Por otra parte, en el caso del modelo para la estimación de la potencia producida por el generador fotovoltaico, se implementa el modelo idealizado de un solo diodo.

Este trabajo de investigación se dividió en dos partes principales: Pronósticos de la radiación solar, y estimación de la potencia producida por el generador fotovoltaico. Los pronósticos de radiación solar, a su vez se subdividen en la parte teórica que trata sobre los procesos internos dentro del Sol, hasta la llegada de la radiación solar a la atmósfera terrestre. Mientras que la otra parte se encarga del análisis y el comportamiento de los datos, es decir, estadística, técnicas avanzadas de minería de datos (pruebas de Dickey – Fuller y pruebas de co-integración) y finalmente los modelos de pronóstico y las pruebas de desempeños para medir la precisión de los mismos.

La segunda parte (estimación de la potencia del generador fotovoltaico), a su vez se subdividió en dos partes principales, el estudio del funcionamiento de la celda fotovoltaica y el modelo idealizado de un solo diodo para determinar la potencia fotovoltaica.

Además, al inicio de este trabajo de investigación se presenta un estudio exhaustivo, de todos los trabajos realizados en este campo.

**PALABRAS CLAVE:** pronóstico, modelos ARIMA, redes neuronales artificiales, modelo fotovoltaico de un sólo diodo, análisis de series de tiempo.

## Abstract

In the present work of investigation the bases for the integration of a model of forecast of the solar radiation are established with a model to calculate the power generated in the photovoltaic systems. In the case of prediction models, three different models were used: the autoregressive stationary integrated model of moving averages (SARIMA), the non-linear autoregressive model of neural networks (NAR) and the non-linear self-regressive model with exogenous inputs of neural networks (NARX). On the other hand, in the case of the model for the estimation of the power produced by the photovoltaic generator, the idealized model of a single diode is implemented.

This research work was divided into two main parts: forecasts of solar radiation, and estimation of the power produced by the photovoltaic generator. The forecasts of solar radiation, in turn, are subdivided into the theoretical part that deals with the internal processes within the Sun, until the arrival of solar radiation to the Earth's atmosphere. While the other part is responsible for the analysis and behavior of the data, ie statistics, advanced data mining techniques (Dickey-Fuller tests and co-integration tests) and finally the forecast models and tests of performances to measure the accuracy of them.

The second part (estimation of the power of the photovoltaic generator), in turn, was subdivided into two main parts, the study of the operation of the photovoltaic cell and the idealized model of a single diode to determine the photovoltaic power.

In addition, at the beginning of this research work an exhaustive study is presented, of all the works highlighted in this field.

## Objetivo General

Desarrollar modelos de predicción, de la potencia generada en equipos fotovoltaicos de gran escala, que sean alimentados por un modelo de pronóstico de la radiación solar, utilizando software especializado en el modelado del sistema fotovoltaico, así como técnicas estadísticas y métodos de inteligencia artificial en la predicción de la radiación.

## Objetivos específicos

- a) Recolectar los datos históricos existentes en México de la radiación solar, a través de los diferentes organismos gubernamentales que se dedican a la medición del recurso solar como es la CFE.
- b) Analizar los datos de radiación solar (minería de datos de radiación solar), con el propósito de eliminar los outliers o datos atípicos que nos puedan generar errores en los modelos de predicción.
- c) Desarrollar modelos de predicción de radiación solar estadísticos (ARX, ARMA, ARIMA) y de computación suave (RNA, AG).
- d) Modelar equipos de radiación solar, para entender su comportamiento y funcionamiento utilizando los resultados de las salidas de los datos pronosticados con los modelos de predicción desarrollados.

## Justificación

Una vez que este proyecto quede concluido, se contará con un mecanismo que nos permita predecir la radiación solar en el lugar seleccionado que permita modelar sistemas fotovoltaicos, con el fin de estimar la factibilidad de instalar éstos, en otros lugares de interés, utilizando la modelación y no la instalación de equipos costosos.

Además, esta investigación se hace necesaria por el hecho de que en la actualidad en México no se han desarrollado trabajos que generen modelos para pronosticar la radiación solar. De la misma manera es limitado el modelado de sistemas fotovoltaicos; lo anterior provoca que se tenga incertidumbre a la hora de utilizar este tipo de plantas para la generación de energía eléctrica, sobre todo a gran escala (30 MW). El desarrollo de modelos completos que incluyan el modelo de los recursos y del sistema, permitirá mayor confiabilidad en la instalación de este tipo de tecnologías en nuestro país.

Los productos finales del presente trabajo, auxiliarán directamente al Sistema Eléctrico Nacional, sobre todo en la toma de decisiones referentes a la instalación de Centrales Eléctricas Fotovoltaicas de gran capacidad. Otros beneficios que van implícitos en el presente trabajo son los que recibirán los consumidores de electricidad y de manera indirecta se atacarán la reducción de problemas de emisión de gases de efecto invernadero, reduciendo la necesidad que existe de utilizar combustibles fósiles para la generación de energía eléctrica.

Teniendo como usuario principal, a la Comisión Federal de Electricidad, y se podrá expandir el área de uso del programa a empresas particulares dedicadas a la producción de energía.

## Hipótesis

Se puede crear un modelo de predicción de la generación de la potencia utilizando como fuente de alimentación la energía solar. Utilizando para su creación técnicas estadísticas y de inteligencia artificial, contando con datos históricos de las mediciones de la radiación solar, además de las características de los equipos fotovoltaicos que se pretenden usar para la generación.

# Capítulo

## I

### Estado del arte

La radiación solar es la fuerza impulsora detrás de varios dispositivos de energía solar, como los sistemas fotovoltaicos para la generación de energía eléctrica y los colectores solares para el calentamiento de agua [1]. Por lo tanto, los modelos de predicción de radiación solar son esenciales para diferentes aplicaciones, desde el diseño de sistemas de energía (PV y sistemas térmicos) hasta aplicaciones meteorológicas [2,3].

La integración de una cantidad grande de plantas fotovoltaicas dentro de la red eléctrica plantea retos técnicos debido a la naturaleza variable del recurso solar. A causa de la tasa exponencial del crecimiento registrado recientemente en el campo de las energías renovables, hay un incremento necesario de modelos más precisos, de pronóstico y predicción de la irradiación solar [4].

Los pronósticos de producción fotovoltaica son especialmente útiles para los operadores de la red para ajustar mejor la generación variable de electricidad en sus programas, despacho y regulación de la potencia. Los horizontes relevantes de predicción pueden ir desde 5 minutos a varios días [4].

Debido a esto, se han realizado grandes esfuerzos con el fin de obtener modelos para el pronóstico de la RSG cada vez más precisos, a continuación, se mencionan algunas de las propuestas y las diferentes técnicas implementadas para la estimación y el pronóstico de la radiación solar.

M. A. Behrang et al. Presentan un trabajo cuyo principal objetivo es predecir la radiación global diaria sobre una superficie horizontal, utilizando diferentes técnicas de redes neuronales y diferentes variables meteorológicas para su entrenamiento, este estudio se llevó a cabo en la ciudad de Dezful, Irán [5]. Mediante el uso de técnicas artificiales Changsong Chen et al. realizan pronósticos de la energía solar, este método requiere como entrada mediciones de energías pasadas y pronósticos meteorológicos de la irradiación solar, humedad relativa y temperatura de los sitios donde se encuentran los sistemas fotovoltaicos (FV). Un mapa auto-organizado (SOM) se entrena para clasificar la ubicación del tipo de clima de 24 h adelante proporcionado por los servicios de datos meteorológicos online [6].

K.S. Reddy et al. Realizan un estudio de modelos de predicción del recurso solar en la india. Estos investigadores comparan los resultados de sus modelos con otros modelos empíricos de regresión. Reportan que las estimaciones de modelos de RNAs tienen buena similitud con los valores actuales y que son superiores a aquellos modelos disponibles [7].

Utilizando técnicas de RNAs Atsu S.S. Dorvlo et al. estiman la RS mediante una primera estimación del índice de claridad, estos autores utilizan modelos de Red de Función de Base Radial y la Perceptrón multicapa, los cuales fueron programados en Matlab [8].

Se propone un modelo de predicción de la RS para determinar el potencial de energía solar en Turquía, por Adnan Sözen et al. Se trata de una Red Neuronal con alimentación directa (feedforward). El gradiente conjugado reducido (SCG), Gradiente conjugado Pola-Ribiere (CGP), y algoritmo de aprendizaje Levenberg-Marquardt (LM) y la función de transferencia logística sigmoidea fueron utilizados en la red [9].

J. Mubiru et al. desarrollan un modelo de predicción utilizando redes neuronales artificiales (ANN). Se utilizan para estimar el promedio mensual diario de la irradiación solar. Duración de la luz solar, temperatura máxima, nubosidad y parámetros de localización: latitud, longitud, altitud son los

datos que alimentan la red. Las salidas de las redes se comparan con métodos empíricos, resultado superiores las predicciones estimadas por las redes [10]. En [11] se presenta un método práctico para el pronóstico de la irradiación solar 24 horas adelante, mediante el uso de RNAs. Este trabajo fue elaborado en Trieste, Italia. La Perceptrón multicapa utilizada consiste de una capa de entrada, una o más capas ocultas y una capa de salida. La RNA se desarrolló en Matlab®, utilizando como algoritmo de entrenamiento Levenberg-Marquardt (LM). Los datos (entrada/salida) fueron normalizados a (-1,1). Ozan Senkal et al. elaboraron RNAs y datos satelitales para estimar la RS en Turquía, utilizaron en sus RNs algoritmos de aprendizaje de Propagación Elástica (RP), Gradiente Conjugado Reducido (SCG), y función de transferencia logística sigmoidea. Datos meteorológicos y geográficos (latitud, longitud, altitud, mes, radiación media difusa y radiación de haz media, radiación solar) se utilizan para entrenar la RNA. La RS ha sido estimada como suma media diaria mensual utilizando datos de Meteosat-6 satellite C3. Se reporta que utilizar ANN y datos satelitales es una manera económica y efectiva para la estimación de la RS y en la construcción de bases de datos de energía solar [12]. Mellit et al. Presentan en su trabajo de investigación un nuevo enfoque para predicción y modelado de datos totales de radiación solar a partir únicamente de la duración media del sol y de la temperatura del aire utilizando un Esquema de inferencia neuro-difuso adaptativo (ANFIS) [13]. Mellit et al. Combinan RNAs y un hardware de predicción de la RS basados en RNA y Lenguajes de Descripción de Hardware (HDLs). Se introduce a resultados preliminares del Campo de Matriz de puertas de programación (FPGA) implementación al modelado solar. Se utilizan datos meteorológicos como: radiación solar, temperatura y duración de la luz solar [14]. A. Ghanbarzadeh et al. pronostican la RSG diaria con datos meteorológicos utilizando RNAs, Los datos utilizados son: temperatura media del aire, la humedad relativa y las horas de luz solar. Las estimaciones de RSG se realizaron utilizando tres combinaciones de conjuntos de datos. Se entrenó una red con alimentación directa (feedforward) para estimar la RSG diaria basada en la duración del día, la temperatura media diaria del aire y la humedad relativa. Las pruebas demostraron que una red con tres entradas, dos capas ocultas, función de transferencia logística sigmoidea (logsig) para ambas capas ocultas, una unidad de salida, función de transferencia lineal (pureline) para la capa de salida, y entrenamiento LM fue suficiente para tal aplicación [15]. Fangping Deng et al. Modelaron la RS utilizando técnicas de RNAs en este trabajo, las RNAs fueron aplicadas para estimar la RSG diaria en China. Se utilizaron datos meteorológicos diarios y el día del año (DOY) para el entrenamiento de la ANN. Se consideraron veinte combinaciones de variables de entrada. Se realizó una comparación con el modelo de regresión Angstrom, los resultados favorecieron a las ANN [16]. Fumitoshi Nomiyama et al desarrollan tres métodos de pronóstico de la RSG utilizando el pronóstico de datos climáticos. Se utilizan para predecir la potencia de salida de los sistemas FV. Dos días adelante, un día adelante y tres horas adelante de RSG, estadística descriptiva, análisis de factor y arboles binarios son las técnicas de análisis de datos y de predicción, utilizada por los autores. La predicción de la RS se obtiene mediante una primera predicción del índice de claridad por tipos de clima [17]. Peng Zhang et al. Proponen Un método para la predicción de la RS basado en la transformada de Wavelet. El método propuesto hace posible extraer las variables de entrada más significantes en cada diferente zona de tiempo – frecuencia [18]. H. Mori et al. Proponen un método de minería de datos para la selección de variables de entrada para los modelos de pronóstico de RSG. El método propuesto hace uso del algoritmo CART (Árboles de Clasificación y Regresión) del método de minería de

datos para seleccionar variables explicativas o de entrada en el modelo de predicción de la RSG [19]. Rui Huang et al. pronostican la generación solar en una micro red a nivel laboratorio, utilizando en la UCLA del centro de investigación de red inteligente (SMERC) como plataforma de prueba. El modelo auto-regresivo de medias móviles (ARMA) y el modelo de persistencia son utilizados para predecir la generación de energía solar futura dentro de las vecindades de UCLA. Los datos históricos de RS se obtienen de SolarAnywhere. Los resultados del pronóstico con el análisis del error indican que el modelo ARMA sobresale en pronósticos a corto y mediano plazo, mientras que la persistencia se desempeña bien únicamente en duraciones a muy corto plazo [20]. Nian Zhang et al. Proponen un estilo Elman basado en redes neuronales recurrentes para predecir la RS a partir de la RS del pasado y la energía solar. Un algoritmo de aprendizaje híbrido incorporando de optimización por enjambre de partículas y algoritmos evolutivos fue implementado [21].

Además de los trabajos citados previamente en lo referente al pronóstico y la estimación de la radiación solar, podemos encontrar trabajos relacionados de una forma más directa con el pronóstico de la radiación solar, como aquellos que utilizan el índice de claridad, como una forma indirecta de predecir la radiación solar [22–26].

Por otro lado, podemos encontrar modelos híbridos, incluidos los que combinan un modelo de predicción con técnicas de agrupación de datos, como el modelo utilizado por [27] que es una combinación de un la técnica de agrupamiento y MLP-k-means NN, [28,29] propusieron modelos híbridos que utilizan diferentes tipos de RNA. Se han desarrollado otros modelos híbridos, por ejemplo, del tipo ARMA / ANN, donde se han utilizado diferentes órdenes ARMA y diferentes tipos de ANN. En estos casos, es importante tener en cuenta que el modelo ARMA se utiliza para la parte lineal, mientras que el modelo ANN se utiliza para la parte no lineal de la serie temporal [28,30]. En algunos casos híbridos, también se usa el índice de claridad, como el enfoque utilizado por [31]. El tercer grupo usa solo técnicas de ANN y actualmente son los modelos más utilizados. [1] compararon los métodos lineales tradicionales versus varias técnicas basadas en la inteligencia artificial. Descubrieron que los modelos de inteligencia artificial predecían las series temporales de radiación solar de manera más efectiva en comparación con los procedimientos convencionales; Se aplicaron FFNN, ER y diferentes funciones de entrenamiento. Otros autores también utilizaron FFNN para la predicción de la radiación solar como [3,32,33]. Algunos trabajos proponen dividir las series temporales anuales en series temporales estacionales [22,31,34] Modelos menos convencionales han sido propuestos por [35,36] a la predicción de radiación solar, como aquellos que usan expresiones polinomiales para la descomposición de las series temporales, el ángulo cenital y la función de transmisión, respectivamente. Un modelo híbrido ARMA-TDNN, la segmentación de datos y el algoritmo de agrupamiento fueron utilizados por [37] para el pronóstico por hora de RSG. [38] utilizaron un modelo acoplado de sistemas autoregresivos y dinámicos y compararon los resultados con un modelo de Lucheroni, finalmente formaron un modelo híbrido. En [39] se llevaron a cabo pruebas de causalidad con el fin de encontrar la relación bidireccional entre las variables meteorológicas; en el caso del estudio presentado, se utilizaron la velocidad del viento, la presión y la temperatura, es importante mencionar que los autores no implementaron ningún modelo de pronóstico.

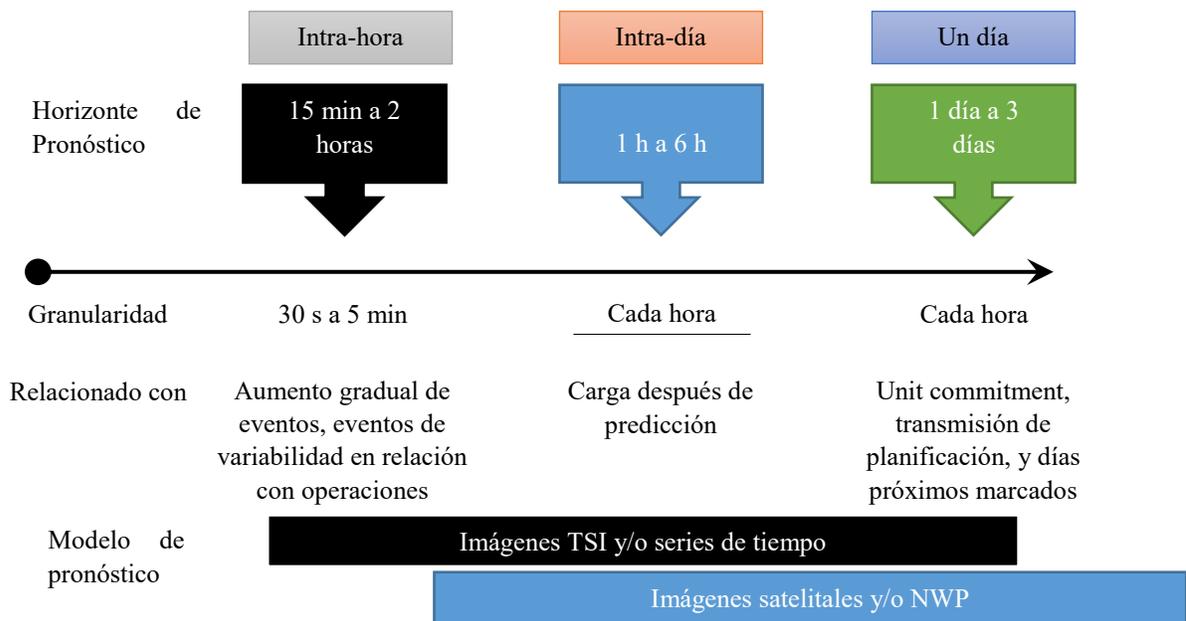
[40] utilizaron el coeficiente de correlación de Pearson como una medida de la correlación lineal entre dos variables para seleccionar las entradas de un modelo de predicción, pero al final los

autores decidieron probar varias combinaciones de entradas para obtener la Irradiación horizontal global (GHI). Por otro lado, en [41] se usó una ANN para pronosticar el GHI y se utilizó la correlación de Pearson para encontrar la correlación entre los datos satelitales y terrestres para obtener las mejores entradas de datos satelitales en los modelos ANN.

Además, [4] revisa varias propuestas y la clasifica en dos grandes grupos principales. El primer grupo se basa en el análisis de las series de tiempo de la irradiación. El segundo grupo utiliza valores pronosticados de un modelo numérico de predicción del tiempo (NWP) e imágenes de nubes.

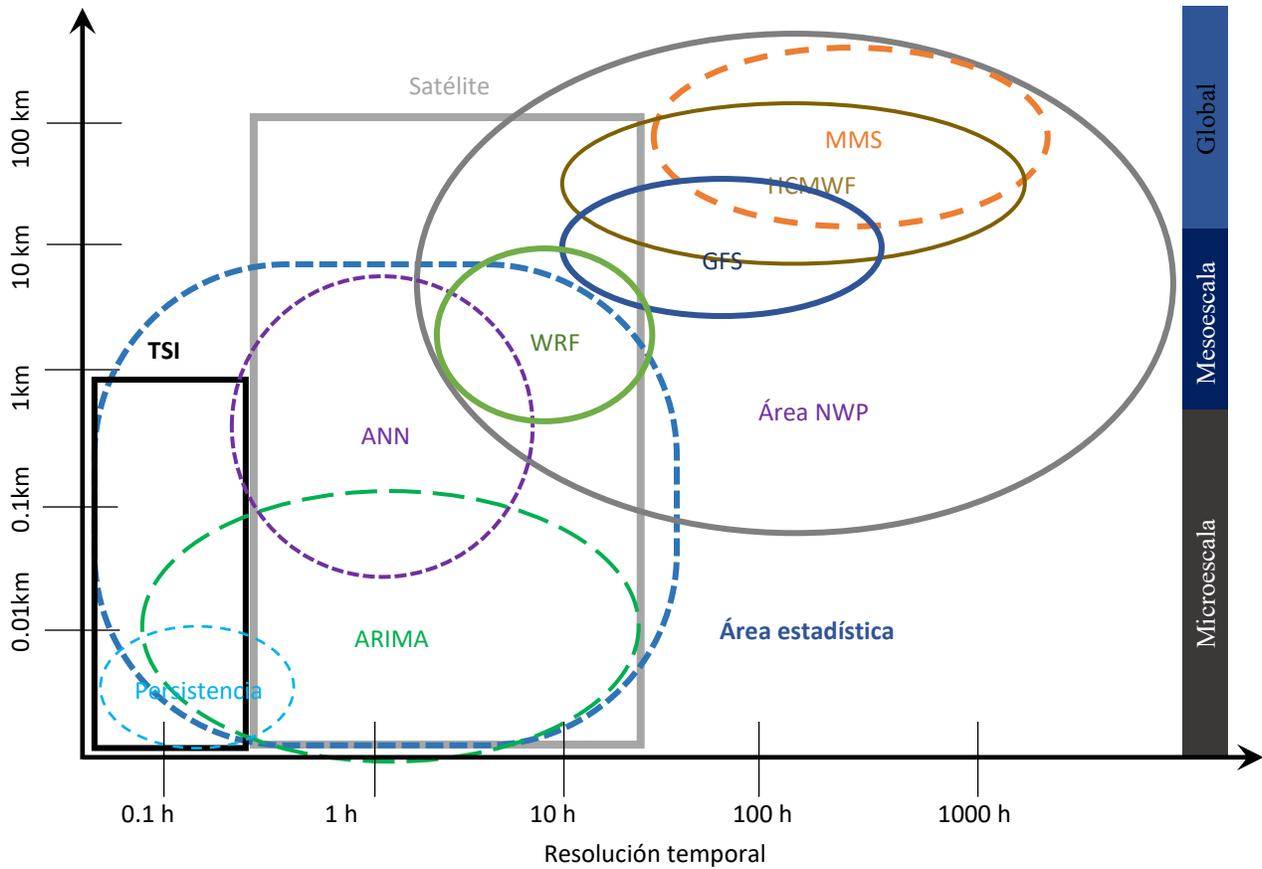
Dentro del primer grupo se tienen:

- Modelo físico
  - Predicción numérica del tiempo
  - Imágenes de nubes
- Modelos estadísticos
  - Persistencia
  - Modelos ARMA
  - Modelos ARIMA
  - Redes Neuronales Artificiales
- Modelo Híbrido



**Figura 1. 1.** Relación entre los horizontes de tiempo, los modelos de predicción y las actividades conexas

En base a la revisión llevada a cabo por [4], la **Figura 1. 1** muestra la relación entre los horizontes de pronóstico, los modelos de pronóstico y actividades relacionadas para los operadores de la red. La clasificación de los modelos de pronóstico basados en la resolución espacial de salida o datos previstos se ilustran en la **Figura 1. 2**.



**Figura 1. 2.** Clasificación de los modelos de predicción basados en resolución espacial de los datos de entrada y resolución temporal

# Capítulo

## 2

### Energía Solar

#### 2.1 Dentro del universo

El Sol es el cuerpo central del sistema solar. Está informalmente designado como una estrella enana amarilla. Pertenece al tipo espectral  $G2V$  donde  $G2$  se refiere a su superficie de temperatura (aproximadamente 5778 K) y  $V$  significa que el Sol es una estrella de secuencia principal y genera su energía por fusión de Hidrógeno en Helio. Otras estrellas de tipo  $GV$  son Alpha Centauri A, Tau Ceti y 51 Pegasi. Cabe señalar que el Sol es blanco, sin embargo, pareciera ser amarillo desde la Tierra debido al efecto de dispersión de Rayleigh a través de la atmósfera [42].

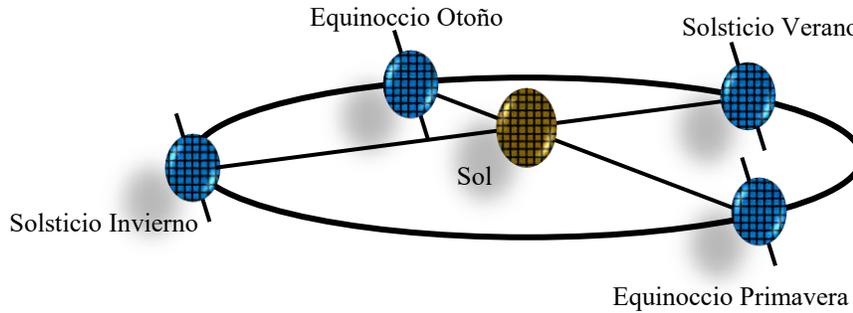
El Sol comparte el movimiento giratorio de la galaxia, además, muestra un mayor movimiento giratorio alrededor de su eje que está inclinado  $7^\circ 15'$  en el plano de la eclíptica. Ya que el Sol tiene que ser considerado como una esfera de gas en lugar de un cuerpo sólido, la velocidad angular no puede ser constante, es menor en los polos y mayor en el ecuador [42].

El Sol es una estrella típica cuyas propiedades han sido estudiadas más profundamente en comparación con otras estrellas por la excepcional cercanía a la Tierra. Por suerte, la física solar coincide con la física de gran parte de nuestras estrellas, galaxias y, con buenas posibilidades, con la de gran parte de las otras galaxias [42].

Debido al giro de la Tierra (que es nuestro punto de observación) en su eje, a la revolución alrededor del sol y finalmente a la distancia de la Tierra al Sol y del Sol a las otras estrellas, experimentamos un movimiento aparente del Sol y de las estrellas. Esto se puede explicar considerando que incluso si las estrellas están siempre en la misma dirección de nosotros todo el día y todo el año (esta es la razón por la cual se los conoce como estrellas fijas), el Sol cambia ligeramente su dirección de nosotros todo el día, desde su ascensión a su puesta durante el año como consecuencia del movimiento de la Tierra a su alrededor [42].

Como se muestra en la **Figura 2. 1**, la eclíptica y el plano ecuatorial tienen dos puntos en común que definen los equinoccios vernales y otoñales [42].

Cuando el Sol está en el equinoccio de primavera, se eleva exactamente hacia el este y se establece exactamente al oeste [42].



**Figura 2. 1.** Trayectoria Sol-Tierra

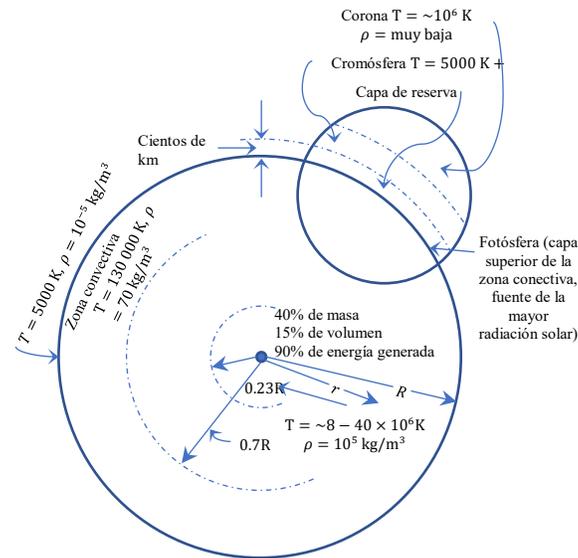
## 2.2. El Sol

El Sol es una esfera de materia gaseosa intensamente caliente con un diámetro de  $1.39 \times 10^9 m$  y se encuentra a una distancia promedio, de  $1.5 \times 10^{11} m$  de la tierra. Como se ve desde la tierra, el sol rota sobre su eje alrededor de una vez cada 4 semanas. No rota como un cuerpo solido; al ecuador le toma alrededor de 27 días y a la región polar le toma sobre 30 días por cada rotación [43].

El sol tiene una temperatura efectiva de cuerpo negro de 5777 K. La temperatura en la región central interior se estima de alrededor de  $8 \times 10^6$  a  $40 \times 10^6$  K y la densidad se estima alrededor de 100 veces la del agua. Se han sugerido varias reacciones de fusión para suministrar la energía radiada por el sol. Una de las más importantes es un proceso en el que el hidrogeno (i.e., cuatro protones) se combinan para formar helio (i.e., un núcleo de helio); la masa del núcleo de helio es menor que la de los cuatro protones, habiendo sido perdida masa en la reacción y convertida en energía [43].

Una estructura esquemática del sol se muestra en la **Figura 2. 2**. Se estima que el 90% de la energía se genera en la región de 0 a  $0.23R$  (donde  $R$  es el radio del sol), que contiene 40% de la masa del sol. A una distancia de  $0.7R$  del centro, la temperatura ha bajado a unos 130 000 K y la densidad disminuye a  $70 kg/m^3$ ; aquí los procesos de convección se empiezan a volver importantes, y la zona de 0.7 a  $1.0R$  se conoce como la **zona convectiva**. Dentro de esta zona la temperatura desciende sobre 5000 K y la densidad es de alrededor  $10^{-5} kg/m^3$  [43].

La superficie del sol parece estar compuesta de gránulos (células de convección irregulares), con dimensiones de 1000 a 3000 km y con un tiempo de vida de unos pocos minutos. Otras características de la superficie del Sol son pequeñas áreas oscuras llamadas poros, que son del mismo orden de magnitud que las células convectivas, y áreas oscuras más grandes llamadas manchas solares, que varían en tamaño. La capa exterior de la zona convectiva se llama fotosfera. El orden de la fotosfera esta nítidamente definida, a pesar de que es de baja densidad (alrededor de  $10^{-4}$  la del aire a nivel del mar). Es esencialmente opaca, como los gases de que se componen están fuertemente ionizados y capaces de absorber un espectro constante de radiación. La fotosfera es la fuente de mayor radiación solar [43].



**Figura 2. 2.** La estructura del Sol [43]

Fuera de la fotosfera esta una atmósfera solar más o menos transparente, observable durante el eclipse solar total o por instrumentos que ocultan el disco solar. Sobre la fotosfera es una capa de gases más fríos de varios cientos de kilómetros de profundidad llamada la **capa inversora** [43]. Fuera de esta, está una capa conocida como la **chromosfera**, con una profundidad de alrededor de 10 000 km. Esta es una capa gaseosa con temperaturas un poco más altas que las de la fotosfera, pero con menor densidad. Todavía más a fuera esta la **corona**, una región de muy baja densidad y de muy alta temperatura ( $10^6$ ) [43].

### 2.2.1 Naturaleza de la radiación solar

#### Definiciones.

**Radiación directa (Haz de Radiación).** La radiación solar recibida del sol sin haber sido dispersada por la atmósfera. (Haz de radiación a menudo se refiere a la radiación solar directa; para evitar confusiones entre los subíndices para directa y difusa, se usa el término haz de radiación.)

**Radiación difusa.** La radiación solar recibida del sol después de que su dirección ha sido cambiada por la dispersión debida a la atmósfera.

**Radiación Solar Total.** La suma del haz de radiación y la radiación solar difusa sobre una superficie. (Las mediciones más comunes de la radiación solar son la radiación total sobre una superficie horizontal, a menudo referido como **radiación global** sobre la superficie.)

**Irradiancia, W/m<sup>2</sup>** La tasa a la que la energía radiante es incidente sobre una superficie por unidad de área o de superficie. El símbolo  $G$  se utiliza para la irradiación solar, con subíndice apropiado para haz, difusa, o de radiación espectral.

**Irradiación o Exposición Radiante, J/m<sup>2</sup>** La energía incidente por unidad de área sobre una superficie, encontrada por integración de la Irradiancia sobre un tiempo especificado, usualmente una hora o un día.

### Definiciones.

**Insolación** es un término aplicado específicamente a irradiación de energía solar. El símbolo  $H$  se utiliza para insolación para un día. El símbolo  $I$  se utiliza para insolación para una hora (u otro periodo si se especifica). El símbolo  $H$  e  $I$  puede representar haz, difusa, o total y puede ser sobre superficies orientadas de cualquier forma.

Los subíndices en  $G$ ,  $H$ , e  $I$  son los siguientes:  $o$  se refiere a radiación sobre la atmósfera de la Tierra, que se refiere a la radiación extraterrestre;  $b$  y  $d$  se refiere a haz de radiación y radiación difusa;  $T$  y  $n$  se refieren a la radiación sobre un plano inclinado y sobre un plano normal en la dirección de propagación. Si ni  $T$  ni  $n$  aparecen, la radiación es sobre un plano horizontal.

**Radiosidad o Emitancia Radiante**,  $W/m^2$  La tasa con la que la energía radiante sale de una superficie por unidad de área al combinar emisión, reflexión y transmisión.

**Potencia de emisión o Auto-Emitancia Radiante**,  $W/m^2$  La tasa a la que la energía radiante sale de una superficie por unidad de área por emisión únicamente.

Cualquiera de estos términos de radiación, excepto insolación, puede aplicarse a cualquier rango de longitud de onda específico (tal como el espectro de energía solar) o a radiación monocromática. La insolación se refiere únicamente a irradiación en espectro de energía solar.

**Hora Solar** Tiempo basado en el movimiento angular aparente del sol a través del cielo con el mediodía solar cuando el sol atraviesa el meridiano del observador.

## 2.3 Geometría del movimiento terrestre

El movimiento terrestre se compone de una traslación alrededor del Sol y un giro sobre su eje. En el movimiento de traslación la Tierra se desplaza alrededor del Sol siguiendo una elipse de baja excentricidad en la que el Sol ocupa uno de los focos. La duración de este movimiento define un año. Este movimiento está contenido en el llamado plano de la eclíptica **Figura 2.1** [44].

Debido a la baja excentricidad de la elipse, la distancia entre el Sol y la Tierra durante el movimiento de traslación es variable. Una ecuación simple para describir esta distancia es la [ecuación 2.1](#) [45]:

$$r = r_0 \left\{ 1 + 0.017 \sin \left[ \frac{2\pi \cdot (d_n - 93)}{365} \right] \right\} \quad (2.1)$$

Siendo  $d_n$  el número de día del año y  $r_0$  es la distancia promedio en este trayecto, denominada unidad astronómica,  $r_0 = 1.496 \times 10^8 \text{ km} = 1 \text{ UA}$  [44].

La corrección debida a la excentricidad de la elipse se calcula con la [ecuación 2.2](#):

$$\epsilon_0 = \left( \frac{r_0}{r} \right)^2 = 1 + 0.033 \cdot \cos \left( \frac{2\pi d_n}{365} \right) \quad (2.2)$$

En el movimiento de giro la Tierra rota sobre si misma alrededor de su eje polar, perpendicular al plano ecuatorial terrestre. Entre el eje polar y el plano de la eclíptica hay un ángulo constante de  $23.45^\circ$ . Sin embargo, el ángulo entre el plano ecuatorial y la línea que une la Tierra y Sol es variable a lo largo del año. Este ángulo variable es la causa de las estaciones, de que el Sol aparezca más alto en los mediodías veraniegos y los días invernales sean más cortos que los de verano. Utilizando

la [ecuación 2.1](#) puede comprobarse sin embargo, que la distancia entre el Sol y la Tierra es mayor en el verano que en el invierno del hemisferio Norte. Así, el efecto debido a la inclinación de los rayos solares es mucho más apreciable en la meteorología que la distancia entre el Sol y la Tierra [44].

Este ángulo se denomina declinación y puede ser calculado de forma aproximada con la [ecuación 2.3](#) (en radianes) y representado en la **Figura 2. 3** [45]. En esta ecuación se supone que la declinación permanece constante a lo largo de un mismo día. Asimismo, el criterio de signos supone considerar positivos los ángulos situados al norte del ecuador terrestre [44].

$$\delta = 23.45^\circ \cdot \sin\left(\frac{2\pi \cdot (d_n + 284)}{365}\right) \quad (2.3)$$

Otros autores han perfeccionado las ecuaciones anteriores. Son destacables las aportaciones de Spencer, Michalsky y Strous [46][47][48]. Como ejemplo, se detalla a continuación la propuesta de Spencer (con el resultado en radianes):

$$X = 2\pi \cdot (d_n - 1)/365 \quad (2.4)$$

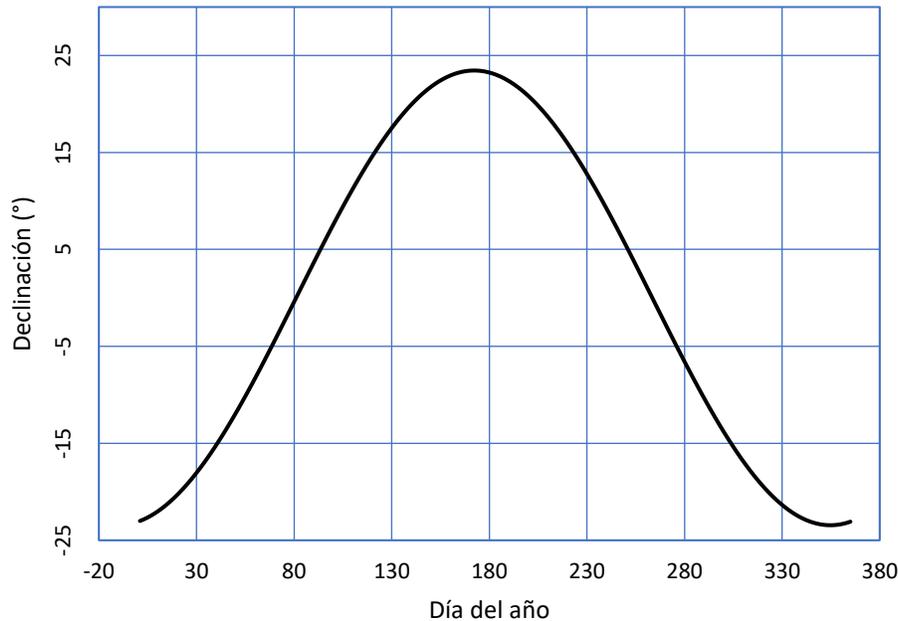
$$\begin{aligned} \delta = & 0.006918 - 0.399912 \cdot \cos(X) + 0.070257 \cdot \sin(X) - 0.006758 \\ & \cdot \cos(2X) + 0.000907 \\ & \cdot \sin(2X) - 0.002697 \cdot \cos(3X) + 0.001480 \cdot \sin(3X) \end{aligned} \quad (2.5)$$

$$\begin{aligned} \epsilon_0 = & 1.000110 + 0.034211 \\ & \cdot \cos(X) + 0.001280 \cdot \sin(X) + 0.000719 \cdot \cos(2X) \\ & + 0.000077 \cdot \sin(2X) \end{aligned} \quad (2.6)$$

El valor de la declinación toma ciertos valores característicos que definen las estaciones y sus fechas de transición. En los equinoccios<sup>1</sup> la declinación es nula, de forma que el Sol amanece y anochece exactamente por el Este y Oeste, respectivamente, siendo equivalentes la duración de día y noche en el solsticio de junio (21-22 junio, día del año 172-173) la declinación toma el valor de  $\delta = 23.45^\circ$ . En el hemisferio Norte es llamado de verano, produciéndose aquí el día más largo del año con el sol amaneciendo por el noreste y anocheciendo por el noroeste. En el solsticio de diciembre (21-22 diciembre, día del año 355-356) la declinación toma el valor de  $\delta = -23.45^\circ$ . En el hemisferio Norte este solsticio es denominado de invierno, ocurriendo el día más corto, con el Sol amaneciendo por el sureste y anocheciendo por el suroeste<sup>2</sup> [44].

<sup>1</sup> En el hemisferio norte el equinoccio de primavera ocurre alrededor del 21-22 marzo (día del año 80-81) y el equinoccio de otoño alrededor del 22-23 septiembre (día del año 265-266).

<sup>2</sup> Estas consideraciones son traducibles a la óptica del hemisferio Sur teniendo en cuenta que en este hemisferio el solsticio de junio es el de invierno, mientras que el de diciembre es el solsticio de verano.



**Figura 2. 3.** Ángulo de declinación

### 2.3.1 Ángulos Tierra – Sol

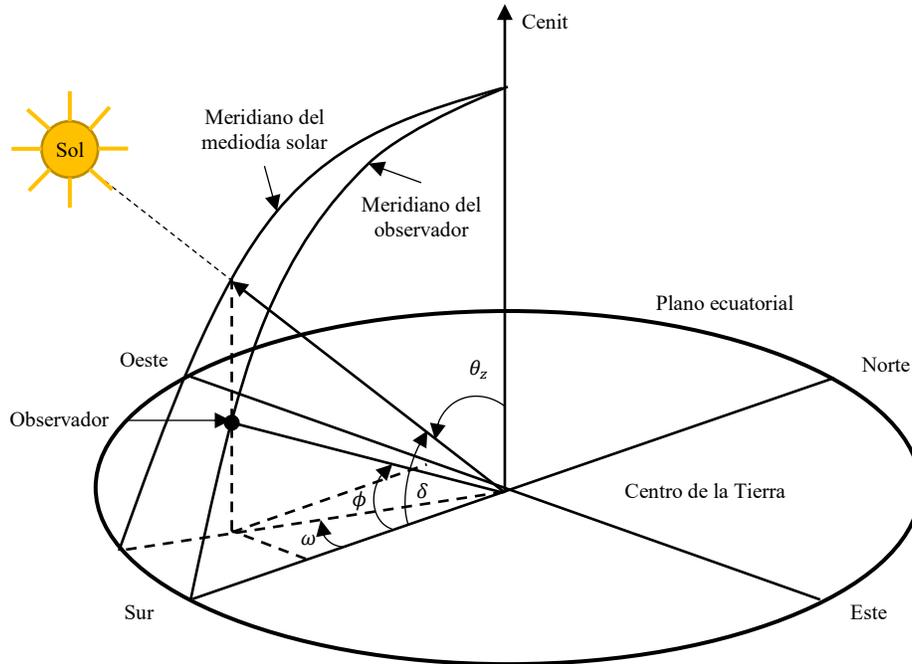
Un mejor entendimiento de los ángulos formados entre la Tierra y el Sol es esencial para estimar la intensidad solar durante todo el año para cualquier superficie en cualquier lugar con una inclinación y orientación deseadas.

**Latitud ( $\phi$ ):** la latitud de un observador (lugar) sobre la superficie de la Tierra es el ángulo formado entre la línea radial unida al observador (lugar) con el centro de la Tierra y su proyección sobre el plano ecuatorial como se muestra en la **Figura 2. 4**. Para un observador en el hemisferio norte, la latitud es positiva, por otro lado, para un observador en el hemisferio sur es negativa.

**Declinación ( $\delta$ ):** la declinación se define como el ángulo entre la línea que une los centros del Sol y la Tierra, que además determina la dirección de los rayos que llegan desde el Sol y su proyección sobre el plano ecuatorial como se muestra en la **Figura 2. 4**. La declinación se debe a la rotación de la Tierra alrededor de un eje, que forma un ángulo de  $66.5^\circ$  con el plano o  $23.5^\circ$  con la rotación normal alrededor del Sol como se muestra en la **Figura 2. 4**. La declinación se obtiene a partir de la siguiente relación [49]:

$$\delta = 23.45 \times \sin \left[ \frac{360}{365} (284 + n) \right] \quad (2.7)$$

A diferencia de la [ecuación 2.3](#), la [ecuación 2.7](#) se obtiene en grados y es la que se implementa en este estudio.



**Figura 2. 4.** Vista de los diferentes ángulos Sol - Tierra

La variación del ángulo de declinación con el  $n$ -ésimo día del año se muestra en la **Figura 2. 3**. El máximo valor de  $\delta$  es de  $23.45^\circ$  (21 de junio), y el mínimo valor de  $\delta$  es de  $-23.45^\circ$  (21 de diciembre) [49].

**Ángulo horario ( $\omega$ ):** este es el ángulo entre las proyecciones de los rayos del Sol (meridiano solar) y la línea que va de sur a norte a través del centro. En otras palabras, el ángulo horario se define como el desplazamiento angular del Sol a partir del meridiano local debido a la rotación de la Tierra alrededor de su propio eje. El ángulo horario correspondiente a 1 hora es  $15^\circ$ . El ángulo horario se puede determinar a partir de la expresión [49]:

$$\omega = (ST - 12) \times 15^\circ \quad (2.8)$$

Donde  $ST$  es la hora local. El ángulo horario total desde el amanecer hasta el ocaso es  $(2\omega_s)$ . El  $\pm\omega_s$  corresponde a un ángulo horario con referencia al amanecer y la puesta de sol, respectivamente [49].

**Cénit ( $\theta_z$ ):** el ángulo entre los rayos del Sol; la línea perpendicular a un plano horizontal se conoce como “ángulo cenital” **Figura 2. 4**.

**Horas de luz (HL):** la duración total de horas del movimiento del Sol desde el amanecer hasta el ocaso. Se define en términos del ángulo horario como

$$HL = \frac{2\omega_s}{15} \quad (2.9)$$

Aquí  $1 \text{ h} = 15^\circ$ .

**Ángulo de incidencia** ( $\theta_i$ ): el ángulo formado entre la normal a la superficie inclinada y la radiación de haz cayendo sobre la superficie inclinada se conoce como “ángulo de incidencia”. En general, el ángulo de incidencia puede ser expresado como [50],

$$\begin{aligned}\cos \theta_i = & (\cos \varphi \cos \beta + \sin \varphi \sin \beta \cos \gamma) \cos \delta \cos \omega \\ & + \cos \delta \sin \omega \sin \beta \sin \gamma \\ & + \sin \delta (\sin \varphi \cos \beta - \cos \varphi \sin \beta \cos \gamma)\end{aligned}\quad (2.10)$$

Para un plano horizontal orientado al sur,  $\gamma = 0$ ,  $\beta = 0$ ,  $\theta_i = \theta_z$  (ángulo cenital)

$$\cos \theta_z = \cos \varphi \cos \delta \cos \omega + \sin \delta \sin \varphi \quad (2.11)$$

De acuerdo a la **Figura 2.4**, el ángulo cenital ( $\theta_z$ ) será de  $90^\circ$  en la mañana como en la tarde antes del amanecer y después del ocaso y  $\omega = \omega_s$ . Por lo tanto, la [ecuación 2.11](#) puede ser escrita de la siguiente manera:

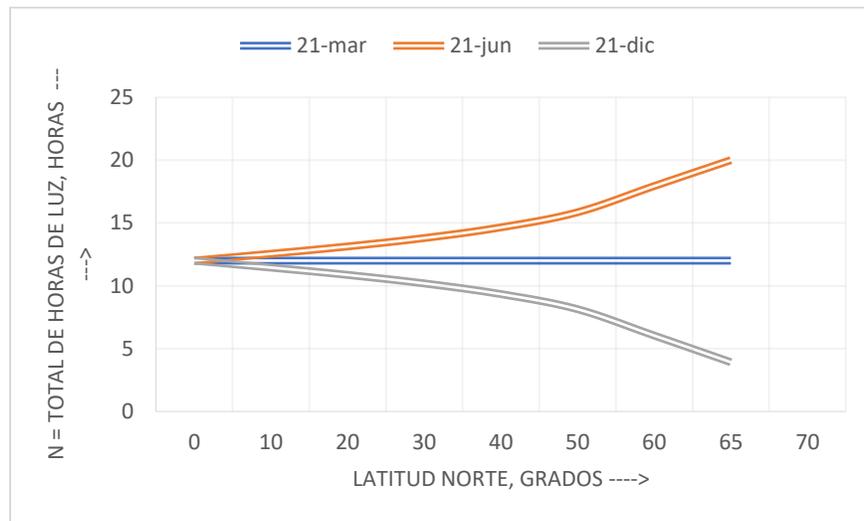
$$\begin{aligned}0 = & \cos \varphi \cos \delta \cos \omega_s + \sin \varphi \sin \delta \\ \omega_s = & \cos^{-1}(-\tan \varphi \tan \delta)\end{aligned}\quad (2.12)$$

El ángulo horario total desde el amanecer hasta la puesta de sol está dado por

$$2\omega_s = 2 \cos^{-1}(-\tan \varphi \tan \delta)$$

Como mencionamos anteriormente,  $1 \text{ h} = 15^\circ$ ; por lo tanto, el número de horas con luz solar ( $HL$ ) está dado por

$$HL = \frac{2}{15} \cos^{-1}(-\tan \varphi \tan \delta) \quad (2.13)$$



**Figura 2. 5.** Variación de las horas de luz solar (HL) con el n-ésimo día del año para diferentes latitudes en el hemisferio norte.

La variación de  $HL$  con el enésimo día del año para diferentes latitudes en el hemisferio norte se muestra en la **Figura 2. 5**.

Ejemplo 2 Determine el número de horas para la ciudad de Morelia para el 22 de diciembre del 2018.

### Solución

Para el presente ejemplo, se tiene que  $\phi = 19.77^\circ$  (Morelia, Mich.); para el 22 de diciembre del 2018,  $n = 356$ , y  $\delta = -23.44^\circ$ . de la ecuación (9) tenemos que

$$\begin{aligned} HL &= \frac{2}{15} \cos^{-1}[-\tan(19.77) \tan(-23.44)] \\ &= \frac{2}{15} \cos^{-1}[-(0.359)(-0.433)] \\ &= \frac{2}{15} \cos^{-1}(0.1558) = 10.8 \text{ hr} \end{aligned}$$

Ejemplo 3 Determinar el ángulo cenital en Morelia a las 2:30 pm el 20 de febrero del 2013.

### Solución

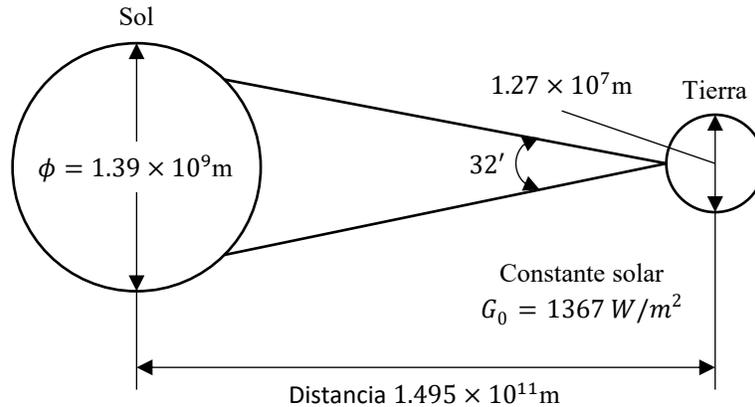
Para el presente caso,  $n = 51$ ;  $\phi = 19.77^\circ$ ;  $\delta = -11.58^\circ$  ecuación 4 y  $\omega = 37.5^\circ$  ecuación 5. De la ecuación 7, tenemos que

$$\begin{aligned} \cos \theta_z &= \cos(19.77) \cos(-11.58) \cos(37.5) + \sin(-11.58) \sin(19.77) \\ &= 0.664 \\ \theta_z &= \cos^{-1}(0.664) = 48.40^\circ \end{aligned}$$

## 2.4 Constante solar

La **Figura 2. 6** muestra esquemáticamente la geometría de la relación Tierra – Sol. La excentricidad de la órbita terrestre es tal que la distancia entre el Sol y la Tierra varía 1.7%. A una distancia de

una unidad astronómica,  $1.495 \times 10^{11}$  m, el promedio de la distancia entre el sol y la tierra, subtendiendo un ángulo de  $32'$ . La radiación emitida por el sol y su relación espacial con la tierra resulta en una intensidad casi fija fuera de la atmósfera terrestre. La constante solar  $B_0$  es la energía del Sol por unidad de tiempo recibida sobre un área unitaria de la superficie perpendicular en la dirección de propagación de la radiación en una distancia promedio entre la tierra y el sol fuera de la atmósfera.



**Figura 2. 6.** Radiación Sol - Tierra

## 2.5 Variación de la Radiación extraterrestre

Dos fuentes de variación en la radiación extraterrestre deben de ser consideradas. La primera es la variación de la radiación emitida por el Sol. Hay informes contradictorios en la literatura sobre las variaciones periódicas de la radiación solar intrínseca. Se ha sugerido que hay pequeñas variaciones (menores que  $\pm 1.5\%$ ) con diferentes periodicidades y variación relacionada a las actividades de las manchas solares.

La variación de la distancia entre la Tierra y el Sol, no obstante, conduce a variaciones de flujo de radiación extraterrestre en el rango de  $\pm 3.3\%$ . La dependencia de la radiación extraterrestre en la época del año se muestra en la **Figura 2. 7**. Una ecuación simple con una precisión adecuada para la mayoría de los cálculos ingenieriles está dada por la [ecuación 2.14a](#). Una ecuación más precisa se proporciona en la forma de la [ecuación 2.14b](#):

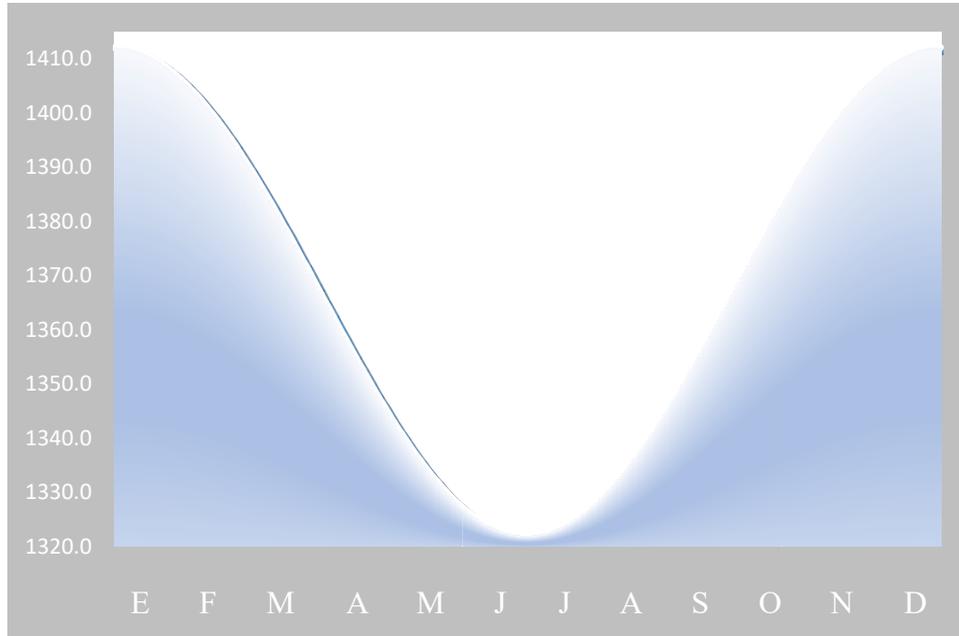
$$I_{ext} = \begin{cases} I_{sc} \left( 1 + 0.033 \cos \frac{360n}{365} \right) & (2.14a) \end{cases}$$

$$\begin{cases} I_{sc} (1.000110 + 0.034221 \cos B + 0.001280 \sin B \\ + 0.000719 \cos 2B + 0.000077 \sin 2B) & (2.14b) \end{cases}$$

Donde  $I_{sc}$  es la radiación incidente extraterrestre en el plano normal a la radiación en el  $n$ -ésimo día del año y  $B$  está dada por

$$B = (n - 1) \frac{360}{365} \quad (2.15)$$

Mientras, que otros autores indican que, Debido a la órbita elíptica del movimiento de la Tierra alrededor del Sol, la distancia Sol-Tierra no es fija y sufre una variación a lo largo del año, alcanzando una variación máxima de 1.7 %. Se puede encontrar gran información al respecto en diferentes fuentes bibliográficas [42][51][52].



**Figura 2. 7.** Variación de la radiación solar extraterrestre con el día del año

**Ejemplo:** Determine la radiación solar extraterrestre el día 22 de mayo de 2018.

Solución. De acuerdo a la **¡Error! No se encuentra el origen de la referencia.**,  $n = 142$ , y tomando  $I_{sc}=1367 W/m^2$ , sustituyendo tenemos que

$$I_{ext} = 1367 \times \left[ 1.0 + 0.033 \cos\left(\frac{360 \times 142}{365}\right) \right]$$

$$I_{ext} = 1332.42 W/m^2$$

**Tabla 2. 1.** Días promedio recomendados de los meses para promedios de radiación del mes

Mes	$N$ para el $i$ -ésimo día del mes	Día del año ( $n$ )
enero	17	$i$
febrero	16	$31 + i$
marzo	16	$59 + i$
abril	15	$90 + i$
mayo	15	$120 + i$
junio	11	$151 + i$
julio	17	$181 + i$
agosto	16	$212 + i$
septiembre	15	$243 + i$

octubre	15	273 + i
noviembre	14	304 + i
diciembre	10	334 + i

Es importante mencionar que la radiación solar en la región extraterrestre tiene cierta dirección. Cuando la radiación pasa a través de los gases de efecto invernadero, hay una atenuación atmosférica. La cantidad de atenuación está gobernada por la masa de aire. La masa de aire es la tasa entre los caminos recorridos por la radiación solar en la atmósfera hasta la masa viajando hasta por la radiación solar si el Sol está en su cenit. Para el mediodía, la tasa es la unidad. Los valores grandes de masa de aire implican mayor atenuación. Una expresión para la masa de aire referida a la **Figura 2. 8** está dada por

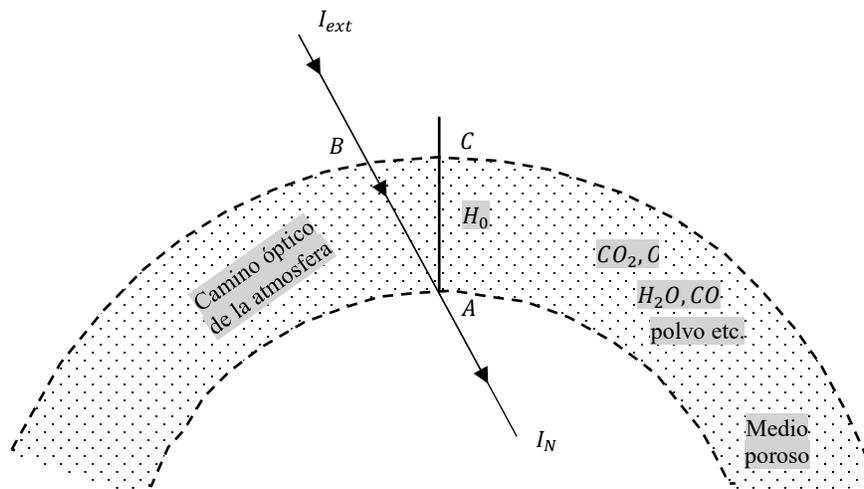
$$MA = \frac{\text{longitud de camino recorrido en la atmósfera}}{\text{profundidad vertical de la atmósfera}} = \frac{AB}{AC} \tag{2.16}$$

$$= \frac{m_0}{H_0} = \sec \theta_z = \frac{1}{\cos \theta_z} \text{ para } \theta_z \leq 0$$

A mediodía,  $\theta_z = 0$ ,  $m = 1$ ; para  $\theta_z = 60^\circ$ ,  $m = 2$  y  $m = 0$  en el exterior de la atmósfera terrestre. Por otro lado, Kasten [53] propuso la modificación de la expresión para la masa de aire de la siguiente forma:

$$m = [\cos \theta_z + 0.15 \times (93.885 - \theta_z)^{-1.253}]^{-1} \tag{2.17}$$

La variación de la masa de aire con la hora del día para la latitud de Chihuahua para diferentes números de días del año se muestra en la **Figura 2. 8**. Se puede observar que los días con luz solar son más cortos y la masa de aire se mayor para el 21 de diciembre comparado con otros días, como se era de esperarse [49].



**Figura 2. 8.** Dirección de los rayos del Sol a través de la atmósfera

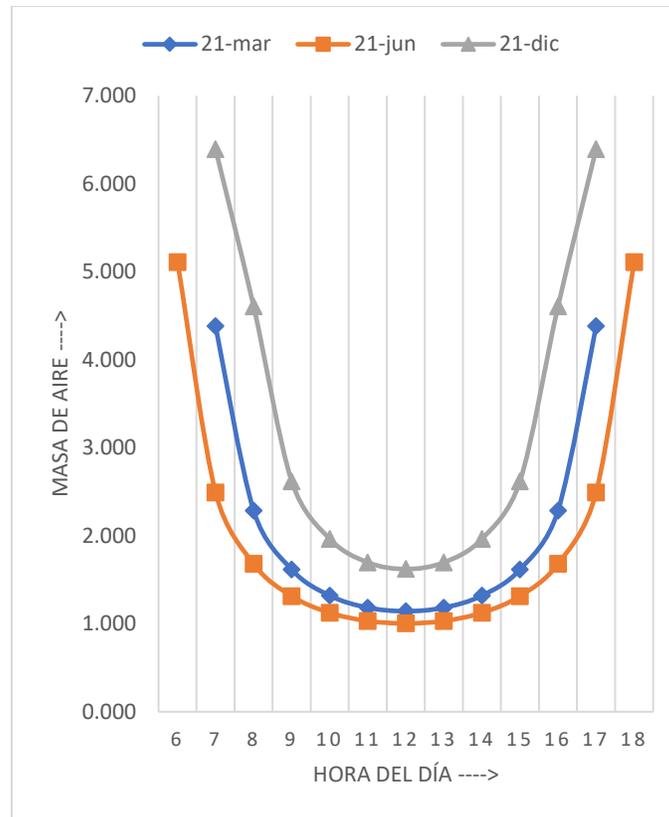


Figura 2. 9. Variación de la masa de aire con la hora del día

## 2.6 Radiación fuera de la atmósfera terrestre.

Aquí, la radiación solar incidente ( $I_0$ ) sobre un plano horizontal en la región extraterrestre fuera de la atmósfera  $W/m^2$ , que es equivalente a estar en ausencia de atmósfera, es una componente de  $I_{ext}$  a lo largo de la normal a la superficie horizontal y puede ser obtenida de las [ecuaciones 2.18](#) y [2.19](#) como sigue:

$$I_0 = I_{extr} \times \cos \theta_z \tag{2.18}$$

o,

$$I_0 = I_{sc} \left[ 1.0 + 0.033 \cos \left( \frac{360n}{365} \right) \right] \cos \theta_z \tag{2.19}$$

Donde  $I_{sc}$  es la constante solar ( $1367 W/m^2$ ), y  $n$  es el enésimo día del año (**¡Error! No se encuentra el origen de la referencia.**). Al sustituir  $\cos \theta_z$  [ecuación 2.11](#) en la [ecuación 2.19](#) obtenemos

$$I_0 = I_{sc} \left[ 1.0 + 0.033 \cos \left( \frac{360n}{365} \right) \right] (\cos \varphi \cos \delta \cos \omega + \sin \delta \sin \varphi) \tag{2.20}$$

La radiación extraterrestre sobre una superficie horizontal para un periodo dado de un ángulo horario  $\omega_1$  a  $\omega_2$  (donde  $\omega_2$  es mayor) puede ser calculada al integrar la [ecuación 2.20](#) para un periodo, y está dada por

$$I_0 = \frac{12 \times 3600}{\pi} I_{sc} \left[ 1.0 + 0.033 \cos \left( \frac{360n}{365} \right) \right] (\cos \varphi \cos \delta (\sin \omega_2 - \sin \omega_1) + \frac{2\pi(\omega_2 - \omega_1)}{360} \sin \delta \sin \varphi) \quad (2.21)$$

La [ecuación 2.21](#), se utiliza para obtener la radiación solar extraterrestre horaria por unidad de área y en MJ, para obtener la radiación extraterrestre horaria en  $W/m^2$  se utiliza la [ecuación 2.22](#) [54].

$$I_{ext} = I_{sc} E_0 (\sin \delta \sin \phi + 0.9972 \cos \delta \cos \phi \cos \omega_{0.5}). \quad (2.22)$$

Donde  $E_0$  es el factor de corrección de excentricidad de la órbita terrestre, la cual se puede obtener a partir de la siguiente expresión:

$$E_0 = 1.000110 + 0.034221 \cos B + 0.001280 \sin B + 0.000719 \cos 2B + 0.000077 \sin 2B \quad (2.23)$$

Dónde  $B$  se puede obtener a partir de la [ecuación 2.15](#).

### 2.6.1 Influencia de la atmósfera terrestre en la radiación solar.

A su paso por la atmósfera, la radiación solar es sometida a una combinación de procesos de reflexión, atenuación y difusión que alteran sus características. La reflexión en las nubes disminuye la radiación incidente en la superficie terrestre mientras que la absorción por vapor de agua, ozono y  $CO_2$  produce una modificación de las características espectrales de la radiación. Además, la dispersión por partículas modifica la distribución espacial de la radiación. Dependiendo del tamaño de la partícula en interacción pueden distinguirse tres fenómenos de difusión [44]:

- Difusión de Rayleigh, donde la longitud de onda es mucho mayor que el tamaño de la partícula. Se produce en las capas altas y es responsable del color azul del cielo.
- Difusión de Mie, donde la longitud de onda es de magnitud similar al tamaño de la partícula. Se produce en las capas bajas de la atmósfera.
- Difusión no selectiva, aquella en la que la longitud de onda es mucho menor que el tamaño de la partícula.

Será de utilidad definir la masa de aire ( $AM$ , *air mass*) como la relación entre el camino recorrido por los rayos directos del Sol a través de la atmósfera hasta la superficie receptora y el que recorrerían en caso de incidencia vertical. Para nuestros efectos es válida la aproximación de la [ecuación 2.24](#), la cual fue definida previamente:

$$AM = 1 / \cos \theta_z \quad (2.24)$$

Para ángulos cenitales mayores, el efecto de la curvatura de la tierra se vuelve significativo y se debe tomar en cuenta<sup>3</sup>.

Para el cálculo de la irradiancia solar que finalmente incide en una superficie arbitraria localizada en la corteza terrestre será útil distinguir tres contribuciones diferentes. Estas contribuciones, comúnmente denominadas componentes, son:

- Radiación Directa,  $B$ : representa la fracción de irradiancia procedente en línea recta del Sol.
- Radiación Difusa,  $D$ : cuantifica la radiación procedente de todo el cielo salvo del Sol, y por tanto incluye todos los rayos dispersados por la atmósfera según los procesos descritos. Es una radiación anisotrópica, cuyo valor depende de la zona celeste de procedencia. Más aún, dado que las propiedades de la atmósfera varían de forma aleatoria con el tiempo, la radiación difusa deberá ser estudiada como un proceso estocástico.
- Radiación del albedo,  $R$  o  $AL$ : es aquella fracción de radiación procedente de la reflexión en el suelo. Habitualmente supone una contribución muy pequeña y en algunos casos puede ser despreciada.

La suma de estas tres componentes constituye la denominada irradiancia global,

$$G = B + D + R \quad (2.25)$$

### 2.6.2 Cálculo de componentes de radiación solar.

Para poder calcular la energía producida por un sistema fotovoltaico o modelar su evolución temporal será necesario contar con información sobre la radiación incidente en el plano de su generador. Cuando se debe predecir la energía producida por el sistema en un tiempo futuro, el problema a resolver consiste en estimar la irradiancia que recibirá a partir de la información disponible sobre el comportamiento de la radiación en aquel lugar. Cuando se trata de un ejercicio de análisis del comportamiento del sistema fotovoltaico, el problema consiste en traducir los valores de radiación disponibles en los correspondientes a la inclinación y orientación del generador. En general, las bases de datos disponibles recogen información sobre la radiación global en el plano horizontal. Muy frecuentemente esta información se almacena en forma de valores de irradiancia diaria, aunque en ocasiones es posible obtener datos de irradiancia de períodos de menor longitud (10 minutos o incluso 1 minuto). Por tanto, para resolver ambos problemas,

<sup>3</sup> Una relación empírica de Kasten y Young (1989) para la masa de aire que funciona para ángulos cenitales cercanos a 90° es

$$m = \frac{\exp(-0.0001184h)}{\cos(\theta_z) + 0.5057(96.080 - \theta_z)^{-1.634}}$$

Donde  $h$  es la altitud del lugar en metros.

llevaremos a cabo diferentes procedimientos de cálculo que obtendrán información sobre la radiación en unas condiciones diferentes a las de partida.

Como será expuesto a continuación, estos procedimientos estimarán la proporción de cada componente en la radiación global. A partir de estas componentes estimadas, se realizarán transformaciones para calcular la radiación directa y difusa en el plano del generador. Veremos que estos procedimientos deben asumir un determinado comportamiento estocástico y una distribución angular de la radiación difusa.

Uno de los primeros intentos de realizar una descripción del comportamiento estocástico de la radiación solar fue llevado a cabo por Liu y Jordan en 1960 [55]. Para caracterizar la radiación solar en un lugar propusieron el índice de claridad,  $K_T$ . Este índice es la relación entre la radiación global y la radiación extra-atmosférica, ambas en el plano horizontal. Según la definición de  $B_0(0)$  dada por la [ecuación 2.26](#), al dividir la radiación global con la extra-atmosférica el índice de claridad no depende de las variaciones debidas al movimiento aparente del sol ni de la localidad. El índice de claridad diario es la relación entre los valores diarios de irradiación:

$$K_{Td} = \frac{G_{d(0)}}{B_{0d(0)}} \quad (2.26)$$

mientras que el índice de claridad mensual es la relación entre las medias mensuales de la irradiación diaria:

$$K_{Tm} = \frac{G_{d,m}(0)}{B_{0d,m}(0)} \quad (2.27)$$

Liu y Jordan propusieron un conjunto de funciones de densidad de probabilidad (FdDPs) para valores diarios del índice de claridad dividiendo la fuente de datos en periodos suficientemente pequeños como para poder considerar que cada periodo está caracterizado por unas características climáticas similares. El periodo mensual parece ser una unidad adecuada. Demostraron que:

1. La forma de la FdDP está muy relacionada con el valor medio del índice de claridad. Dicho en otras palabras, la distribución de valores que presenta la radiación solar durante un periodo está determinada por el valor promedio de la radiación durante ese periodo. Por ejemplo, conocer la media mensual de la radiación solar diaria en un determinado lugar permite saber cómo se comportará la radiación diaria durante ese mes [56].
2. Sólo es significativa la autocorrelación con un día de retraso. Es decir, el índice de claridad para un día concreto sólo está influido por el índice de claridad del día anterior.

Relacionaron este índice de claridad con la fracción de radiación difusa,  $F_D$ , calculada como la relación entre la irradiación difusa y la irradiación global, ambas en el plano horizontal:

$$F_D = \frac{D(0)}{G(0)} \quad (2.28)$$

Cuando el cielo es claro ( $K_T$  tiende a 1) la nubosidad es poco densa y por tanto la fracción de difusa debe ser baja ( $F_D$  tiende a 0). Los cielos nubosos estarán caracterizados por una fracción de difusa elevada, pero por índices de claridad bajos. Este análisis intuitivo fue aplicado al análisis de multitud de medidas de estaciones terrestres que recogían simultáneamente valores de radiación difusa y radiación global, y se comprobó la existencia de una correlación negativa entre ambos parámetros. Más aún, dada la definición del índice de claridad, la expresión de la correlación para medias mensuales se hace independiente de la localidad y adquiere validez casi universal. Es de uso frecuente la correlación para índices mensuales propuesta por [57]:

$$F_{Dm} = 1 - 1.13 \cdot K_{Tm} \quad (2.29)$$

Por ejemplo, un lugar que recibe en el plano horizontal  $3150 \frac{Wh}{m^2}$  de media mensual de irradiación global diaria en un mes que corresponde a media mensual de irradiación extraterrestre diaria de  $4320 \frac{Wh}{m^2}$  tendrá, en ese mes, un índice de claridad mensual  $K_{Tm} = \frac{3150}{4320} = 0.73$  y, según la correlación de Page, una fracción de difusa  $F_{Dm} = 1 - 1.13 \cdot 0.73 = 0.175$ . Por tanto, la media mensual de radiación difusa diaria será  $D_{d,m}(0) = 0.175 \cdot 3150 = 551.6 Wh/m^2$  y la radiación directa será  $B_{d,m}(0) = 3150 - 551.6 = 2598.4 Wh/m^2$ .

De la misma forma, se pueden establecer correlaciones entre valores diarios del índice de claridad y la fracción de difusa. Sin embargo, al disminuir la escala temporal la dispersión de valores aumenta sensiblemente y así el error asociado a las regresiones propuestas. El funcionamiento de un sistema fotovoltaico está determinado en primer lugar por la radiación incidente, aunque existen otros factores de segundo orden que no se pueden despreciar, principalmente la temperatura.

Cuando sea necesario el cálculo de la radiación difusa en un día determinado, es recomendable la correlación propuesta por Collares Pereira y Rabl [58]:

$$F_{Dd} = \begin{cases} 0.99 & K_{Td} \leq 0.17 \\ 1.88 - 2.272 \cdot K_{Td} + 9.473 \cdot K_{Td}^2 - 21.856 \cdot K_{Td}^3 + 14.648 \cdot K_{Td}^4 & K_{Td} \geq 0.17 \end{cases} \quad (2.30)$$

## 2.7 Irradiación solar bajo condiciones de cielo despejado

El  $k_T$  es el cociente de la irradiación extraterrestre que llega a la superficie terrestre en condiciones de cielo despejado  $I_{clr}$ , dividida por la irradiación solar registrada  $G_I$ . Se sabe que, al pasar de la región extraterrestre a la región terrestre a través de la atmósfera terrestre, sufre pérdidas de dispersión y absorción atmosférica [49]. Después de la absorción por la atmósfera, la tasa de flujo solar normal (radiación solar / irradiancia normal) que llega a la superficie de la Tierra viene determinada por:

$$I_{clr} = I_{ext} \times \exp \left[ \frac{-T_R}{(0.9 + 9.4 \cos \theta_z)} \right], \quad (2)$$

Dónde  $T_R$  es el factor de turbiedad y depende de las características geográficas de cada lugar, y  $\cos \theta_z$  se puede obtener a partir de la [ecuación 2.11](#)

# Capítulo

## 3

### La celda solar

#### 3.1 El efecto fotovoltaico

La transformación de la energía radiante proveniente del Sol en energía eléctrica implica el estudio de las interacciones de las ondas electromagnéticas con la materia. Este mecanismo puede ser entendido comenzando a partir del efecto fotoeléctrico en el que los electrones son emitidos desde un material cuando este se expone a la radiación electromagnética [42].

En particular, se observó que (usando luz visible para metales alcalinos, casi ultravioleta para otros metales y ultravioleta extremo para no metales) la energía de los electrones emitidos aumentaba con la frecuencia y no dependía de la intensidad de la radiación [42].

Este efecto fue observado por primera vez por Heinrich Hertz en 1887 y durante varios años aparentemente contrastaba con la teoría ondulatoria de la luz de James Clerk Maxwell; de acuerdo con esta teoría, la energía del electrón sería proporcional a la intensidad de la radiación [42].

Se observaron los siguientes resultados experimentales principales para el material dado:

1. La velocidad a la que se expulsan los fotoelectrones es directamente proporcional a la intensidad de la luz incidente;
2. Existe una frecuencia de umbral, debajo de la cual no se emiten fotoelectrones;
3. Por encima de la frecuencia de umbral, si se aumenta la intensidad de la luz, también aumenta el número de electrones emitidos, pero su energía máxima no varía; además, una intensidad de luz incidente muy baja, con una frecuencia mayor que el umbral, es capaz de extraer electrones;
4. Por encima de la frecuencia de umbral, si se aumenta la frecuencia de la luz incidente, también se aumenta la energía máxima de los fotoelectrones.

Albert Einstein teorizó, en 1905, que la luz está compuesta de cuantos discretos, ahora llamados fotones, y que la energía de un cuanto de luz está dada por el producto de la frecuencia de la onda correspondiente multiplicada por una constante, más tarde llamada constante de Planck [42].

$$E = h\nu \quad (3.1)$$

donde  $E$  es la energía de un solo cuanto de luz,  $h = 6.62 \times 10^{-34} \text{ J} \cdot \text{s}$  es la constante de Planck y  $\nu$  es la frecuencia de la onda electromagnética (fotón incidente). Por este descubrimiento, Einstein obtuvo el Premio Nobel de Física en 1921 y comenzó la revolución cuántica en física [42].

Si un fotón tiene suficiente energía, puede dar su energía a un electrón y es expulsado.

La energía cinética máxima  $K_{m\acute{a}x}$  de un electrón eyectado viene dada por:

$$K_{m\acute{a}x} = h\nu - \varphi \tag{3.2}$$

donde  $\varphi = hf_0$  es la funci3n de trabajo (a veces denominada  $W$ ), que es la energa m3nima requerida para eliminar un electr3n deslocalizado de la superficie de cualquier metal dado. Se deduce que la frecuencia  $\nu$  del fot3n incidente debe ser mayor que  $f_0$  para extraer electrones.

La configuraci3n experimental para verificar el efecto fotoel3ctrico se esquematiza en la **Figura 3.1**.

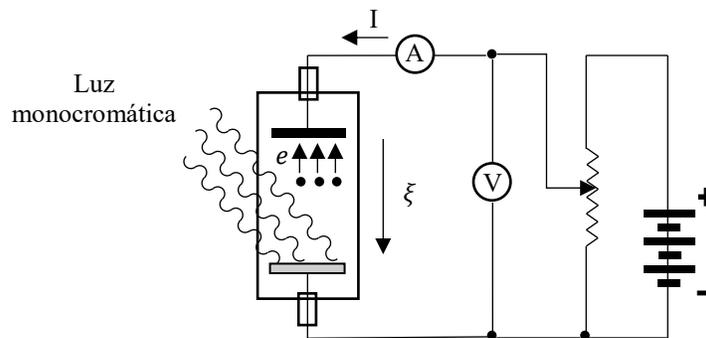
**1.** El haz de luz golpea el material que se encuentra dentro de una bombilla. Dos placas, sujetas a un potencial variable, generan un campo el3ctrico  $\xi$  que puede romper los fotoelectrones. El galvan3metro mide la corriente generada por fotoelectrones.

Cuando la frecuencia del haz incidente es mayor que la frecuencia umbral, aplicando un potencial tal que  $K_{m\acute{a}x} = eV$ , la corriente medida es nula.

Cabe se1alarse que es conveniente medir la energa en electr3n voltio [eV] en lugar de Joule ( $1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$ ).

Es f3cil calcular que la energa asociada con un fot3n emitido por el Sol que tiene  $\lambda = 0.5 \mu\text{m}$  es igual a:

$$E|_{\lambda=0.5\mu\text{m}} = h\nu = h \frac{c}{\lambda} = 6.62 \cdot 10^{-34} \frac{3 \cdot 10^8}{0.5 \cdot 10^{-6}} = 3.972 \cdot 10^{-19} \text{ J} = 2.48 \text{ eV}$$



**Figura 3.1.** Configuraci3n experimental para la evaluaci3n del efecto fotoel3ctrico

### 3.1.1 Metales, semiconductores, aislantes

En un solo 3tomo aislado, los niveles de energa de los electrones son discretos. Para el 3tomo de hidr3geno, el modelo de Bohr da:

$$E_H = \frac{-m_0 q^4}{8 \epsilon_0^2 h^2} \frac{1}{n^2} \tag{3.3}$$

donde  $m_0$  es la masa del electr3n libre y  $q$  su carga,  $\epsilon_0^2$  es la permitividad del espacio libre y  $n$  es un entero positivo conocido como n3mero cu3ntico principal. El nivel fundamental corresponde a  $n = 1$  y la energa relacionada es  $E_{H(n=1)} = -13.6 \text{ eV}$ .

<sup>4</sup> El valor negativo de esta energa corresponde a un electr3n vinculado

Si los átomos de  $N$  interactúan (por ejemplo, en un cristal),  $N$  niveles externos tienen energía solo ligeramente diferente y la energía térmica permite que los electrones pasen de un nivel a otro (la energía correspondiente a  $T = 300\text{ K}$  es  $k_T \approx 0.026\text{ eV}$ ). Los niveles de energía resultantes se agrupan en bandas. Dos bandas principales son reconocibles: banda de conducción y banda de valencia. Estas dos bandas están separadas por una región prohibida que se caracteriza por un valor de energía  $E_g$ . Este valor hace la diferencia entre aisladores, conductores y semiconductores.

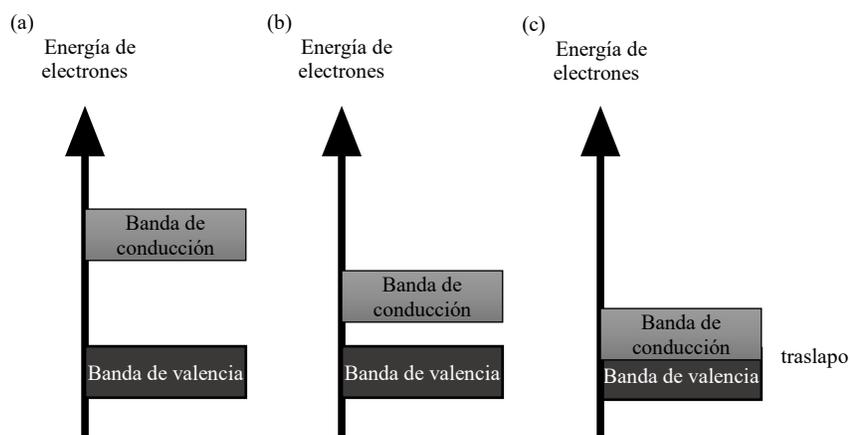
En un aislador, la banda prohibida tiene una amplia energía (por ejemplo,  $E_g = 9\text{ eV}$  para  $\text{SiO}_2$ ), ni la energía térmica ni un campo eléctrico puede elevar la energía de un electrón para enviarlo a la banda de conducción. Debido a la ausencia de electrones libres para la conducción, el material se comporta como un aislante.

Por el contrario, en un conductor, la banda de conducción está parcialmente superpuesta a la banda de valencia. Como consecuencia, hay muchos electrones disponibles para la conducción y un campo eléctrico puede darles suficiente energía para realizar la conducción.

En un semiconductor, las dos bandas están separadas pero la energía de la banda prohibida es baja ( $E_g = 1.12\text{ eV}$  para Si en  $T = 300\text{ K}$ ) y es fácil dar energía a un electrón para entrar en la banda de conducción. En este caso, el orificio en la banda de valencia contribuye tanto a la conducción como al electrón en la banda de conducción **Figura 3. 2**. La amplitud de banda prohibida varía con la temperatura, para el Si la amplitud es:

$$E_g(T) = 1.17 - \frac{(4.73 \cdot 10^{-4})T^2}{T + 636} \quad (3.4)$$

El coeficiente de temperatura es negativo, significa que la amplitud prohibida de la banda disminuye con la temperatura.



**Figura 3. 2.** Representación esquemática de la estructura de banda para: a aislante, b semiconductor, c conductor

### 3.1.2 Dentro de la estructura de banda de un semiconductor

El movimiento de los electrones dentro de un cristal está influenciado por la estructura cristalina periódica del semiconductor que determina un campo potencial periódico debido al núcleo del átomo y los electrones del núcleo fuertemente unidos.

El comportamiento dinámico se obtiene al resolver la ecuación de Schrödinger.

$$\Delta^2\psi + \frac{2m_0}{\left(\frac{h}{2\pi}\right)^2} [E - U(\vec{r})]\psi = 0 \quad (3.5)$$

donde  $m_0$  es la masa libre del electrón,  $E$  es la energía del electrón, y  $U(\vec{r})$  es el potencial periódico dentro del semiconductor. La solución de la ecuación de Schrödinger define la estructura de la banda, la energía permitida y el momento cristalino  $p = \hbar k$  donde  $k$  es el vector de onda y  $\hbar = h/2\pi$ .

La relación entre la energía y el impulso cerca del mínimo de la banda de conducción y cerca del máximo de la banda de valencia es casi parabólica; como consecuencia, la masa efectiva se define como la inversa de la segunda derivada de la energía con respecto al momentum<sup>5</sup>:

$$m^* \equiv \left[ \frac{d^2E}{dp^2} \right]^{-1} \quad (3.6)$$

Esta definición permite que una partícula se considere como una partícula clásica con una masa efectiva  $m^*$ . La masa efectiva  $m^*$  depende de la banda. Cerca de la parte superior de la banda de valencia este valor es negativo. Estos lugares suelen estar vacíos debido a los electrones térmicamente excitados que se encuentran en la banda de conducción.

Es más fácil considerar estos estados como ocupados por otro tipo de partículas, dichos agujeros que se comportan como partículas libres clásicas con una masa efectiva positiva  $m_p^*$ , mientras que los electrones se consideran partículas libres clásicas con masa efectiva  $m_p^*$ .

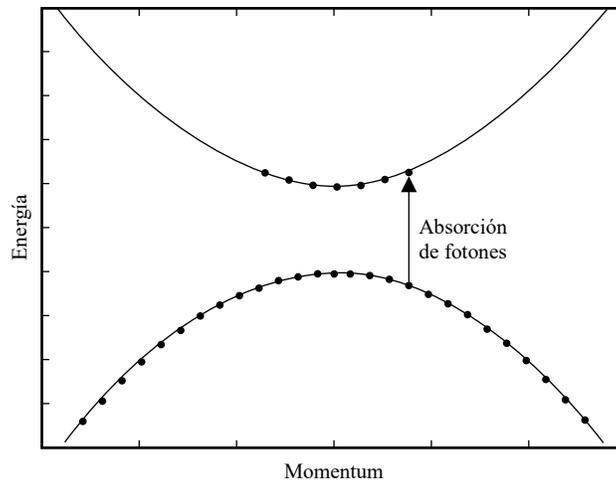
La **Figura 3.3** describe un diagrama simplificado de bandas de energía para un semiconductor en  $T > 0$  en diagrama de energía versus diagrama de momento.

Cabe señalar que el mínimo de la banda de conducción se produce para los mismos valores de momentum del máximo de la banda de valencia. Como consecuencia, un electrón puede hacer una transición manteniendo constante el impulso. Este tipo de semiconductor es un semiconductor de banda prohibida directa.

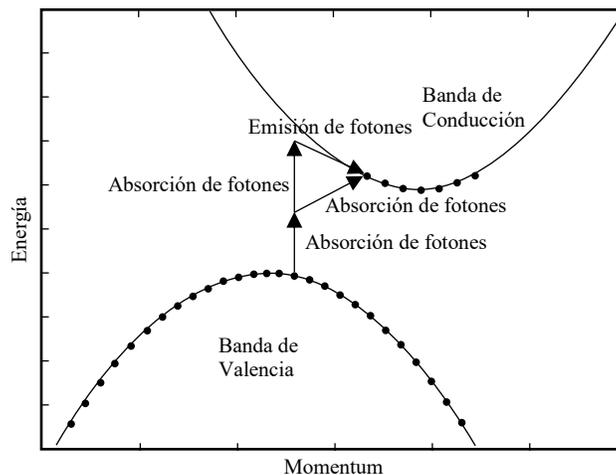
El comportamiento de un semiconductor de banda prohibida indirecta y el diagrama de banda de energía se ilustra en la **Figura 3.4**.

La diferencia entre estos dos tipos de semiconductores es importante para el estudio de la absorción y emisión de luz en un semiconductor.

<sup>5</sup> Debe tenerse en cuenta que  $E = p^2/2m$



**Figura 3. 3.** Energía versus representación del momentum de la estructura de la banda de energía para un semiconductor de banda prohibida directa



**Figura 3. 4.** Energía versus la representación de momentum de la estructura de banda de energía para el semiconductor de banda prohibida indirecto

### 3.1.3 Absorción de la luz

Como se explica en la [sección 3.1.1](#), la energía radiada interactúa con la materia, incluidos los semiconductores, como fotones, cuya energía es  $E = h\nu$ , y el impulso  $p_\lambda = h/\lambda$ .

La excitación de un electrón de la banda de valencia a la banda de conducción se denomina absorción fundamental y, como consecuencia, aparece un orificio en la banda de valencia.

Tanto la energía total como el impulso deben conservarse; en particular, para semiconductores de banda prohibida directa (GaAs, GaInP, CdTe y CU (InGa) Se<sub>2</sub>) puede ocurrir una transición que permanezca constante el momentum del fotón como se muestra en la **Figura 3. 3**. El momentum de cristal es igual a  $p_\lambda = h/l$  donde  $l$  es la constante de la red y es más grande que el momento del

fotón. Siendo la longitud de onda de la luz solar del orden de  $10^{-4}$  cm y la retícula constante de  $10^{-8}$  cm, se puede suponer que la ley de conservación solo se puede aplicar al momentum del fotón.

La probabilidad de una transición inducida de un nivel  $E_1$  en la banda de valencia a un nivel  $E_2$  en la banda de conducción para un fotón con energía  $E_2 - E_1 = h\nu$  viene dado por un coeficiente  $\alpha$  que depende de la diferencia entre la energía del fotón y la banda prohibida.

$$\alpha(h\nu) \approx (h\nu - E_g)^{\frac{1}{2}} \quad (3.7)$$

Algunos semiconductores permiten únicamente transiciones con  $p \neq 0$ , en tales casos:

$$\alpha(h\nu) = \frac{1}{h\nu} (h\nu - E_g)^{\frac{3}{2}} \quad (3.8)$$

En semiconductores de banda prohibida indirectos, como Si y Ge, el máximo de la banda de valencia y el mínimo de banda de conducción se producen para diferentes valores de la cantidad de movimiento, tal como se esboza en la **Figura 3. 4**. La conservación del impulso implica en este caso la emisión o la absorción de un fonón<sup>6</sup>. En particular, si la energía del fotón es mayor que la diferencia entre el nivel de energía del electrón inicial en la banda de valencia y el nivel final en la banda de conducción, un fonón es emitido. Por el contrario, si la energía del fotón es menor que la diferencia entre el nivel de energía del electrón inicial en la banda de valencia y el nivel final en la banda de conducción, se absorbe un fonón.

El coeficiente de absorción es diferente según el fenómeno de absorción ( $\alpha_a(h\nu)$ ) o emisión ( $\alpha_e(h\nu)$ ).

$$\alpha_a(h\nu) \propto \frac{(h\nu - E_g + E_{ph})^2}{e^{E_{ph}/kT} - 1} \quad (3.9)$$

$$\alpha_e(h\nu) \propto \frac{(h\nu - E_g + E_{ph})^2}{1 - e^{E_{ph}/kT}} \quad (3.10)$$

donde  $E_{ph}$  es la energía fonónica. Cabe señalar que, para semiconductores de banda prohibida indirecta, la absorción de un fotón depende de la disponibilidad de estados de energía, y en los fonones absorbidos / emitidos también. Esto hace que el coeficiente de absorción para la transición indirecta sea más pequeño que el correspondiente para la transición directa. Como resultado, la luz puede penetrar más dentro de un semiconductor de banda prohibida indirecta.

<sup>6</sup> El fonón se considera como una representación de partículas de una vibración reticular en el semiconductor, es una partícula de baja energía con un impulso relativamente alto

### 3.1.4 Estados permitidos para huecos y electrones

Mediante la ecuación de Schrödinger, es posible obtener la densidad de estados para electrones y huecos. La influencia del potencial periódico se incorpora a la masa efectiva. La densidad de los estados disponibles para el volumen de la unidad y la energía da los niveles de energía permisibles que podrían ser ocupados si la partícula correspondiente tiene suficiente energía.

En la banda de conducción  $E > E_C$ , la densidad de estados está dada por:

$$g_C(E) = \frac{m_n^* \sqrt{2m_n^*(E - E_C)}}{\pi^2 \hbar^3} \quad [\text{cm}^{-3}\text{eV}^{-1}] \quad (3.11)$$

Y en la banda de valencia ( $E < E_V$ ) la densidad de estados está dada por:

$$g_V(E) = \frac{m_p^* \sqrt{2m_p^*(E_V - E)}}{\pi^2 \hbar^3} \quad [\text{cm}^{-3}\text{eV}^{-1}] \quad (3.12)$$

### 3.1.5 Distribución de energía para huecos y electrones

Teniendo en cuenta un semiconductor en equilibrio térmico, sin generación de portador y aislado externamente, la probabilidad de que un estado esté ocupado por un electrón con energía  $E$  viene dada por la función de Fermi.

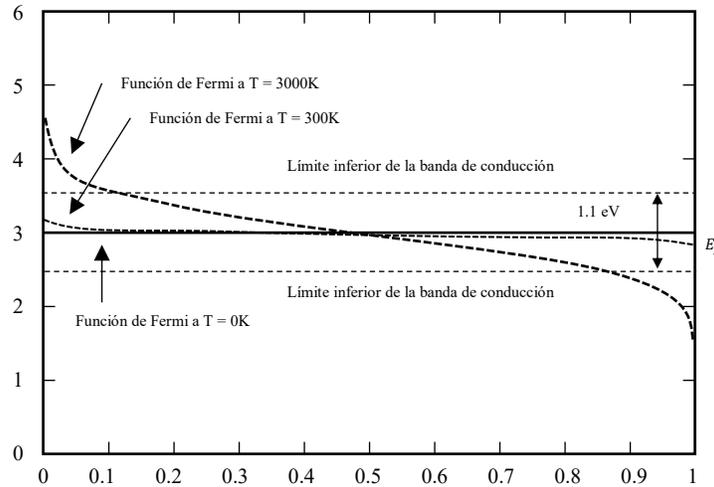
$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}} \quad (3.13)$$

donde  $E_F$  es la energía de Fermi. Este valor depende del material.

En  $T = 0$ , la función de Fermi es una función de paso. Todos los electrones tienen energía por debajo o igual a  $E_F$ . Cuando la temperatura aumenta hay una probabilidad no nula de que los electrones puedan tener energía por encima de  $E_F$ . En la **Figura 3.5**, la función de Fermi se traza para  $T = 0\text{K}$ ,  $T = 300\text{K}$  y  $T = 3000\text{K}$ ; El nivel de Fermi se establece en  $3\text{eV}$  y el intervalo de banda es  $1.1\text{eV}$  como en el Silicio. Cabe señalar que, a temperatura ambiente, el número de electrones con energía mayor que el límite inferior de la banda de conducción es insignificante.

En la banda de conducción, un estado está ocupado si hay un estado disponible y si un electrón tiene suficiente energía. Como consecuencia, para obtener la densidad de electrones de equilibrio es necesario multiplicar la correspondiente densidad de estado [ecuación 3.11](#) por la probabilidad de que el electrón tenga la energía correspondiente [ecuación 3.13](#) e integrarla desde el nivel mínimo que se va a ocupar ( $E_C$ ) hasta el infinito.

$$n_0 = \int_{E_C}^{\infty} g_C(E) f(E) dE = 2 \left( \frac{2\pi m_n^* kT}{h^2} \right)^{3/2} e^{-(E_C - E_F)/kT} \quad (3.14)$$



**Figura 3. 5.** Nivel de Fermi y densidad de estado

dónde la siguiente suposición se puede hacer:

$$N_C = 2 \left( \frac{2\pi m_n^* kT}{h^2} \right)^{3/2} \tag{3.15}$$

$N_C$  es la densidad efectiva de estado en la banda de conducción. De la misma manera, la concentración de huecos en equilibrio está dada por:

$$p_0 = \int_{-\infty}^{E_V} g_V(E)[1 - F(E)]dE = 2 \left( \frac{2\pi m_p^* kT}{h^2} \right)^{3/2} e^{-(E_V - E_F)/kT} \tag{3.16}$$

Y se puede hacer la siguiente suposición:

$$N_V = 2 \left( \frac{2\pi m_p^* kT}{h^2} \right)^{3/2} \tag{3.17}$$

Periodo		13	14	15	16 ← Grupo
↓	2	5 <b>B</b> <small>2s<sup>2</sup>2p<sup>1</sup></small>	6 <b>C</b> <small>2s<sup>2</sup>2p<sup>2</sup></small>	7 <b>N</b> <small>2s<sup>2</sup>2p<sup>3</sup></small>	8 <b>O</b> <small>2s<sup>2</sup>2p<sup>4</sup></small>
	3	13 <b>Al</b> <small>3s<sup>2</sup>3p<sup>1</sup></small>	14 <b>Si</b> <small>3s<sup>2</sup>3p<sup>2</sup></small>	15 <b>P</b> <small>3s<sup>2</sup>3p<sup>3</sup></small>	16 <b>S</b> <small>3s<sup>2</sup>3p<sup>4</sup></small>
	4	31 <b>Ga</b> <small>3d<sup>10</sup>4s<sup>2</sup>3p<sup>1</sup></small>	32 <b>Ge</b> <small>3d<sup>10</sup>4s<sup>2</sup>3p<sup>2</sup></small>	33 <b>As</b> <small>3d<sup>10</sup>4s<sup>2</sup>3p<sup>3</sup></small>	5 <b>Se</b> <small>3d<sup>10</sup>4s<sup>2</sup>3p<sup>4</sup></small>
	5	49 <b>In</b> <small>4d<sup>10</sup>5s<sup>2</sup>5p<sup>1</sup></small>	50 <b>Sn</b> <small>4d<sup>10</sup>5s<sup>2</sup>5p<sup>2</sup></small>	51 <b>Sb</b> <small>4d<sup>10</sup>5s<sup>2</sup>5p<sup>3</sup></small>	52 <b>Te</b> <small>4d<sup>10</sup>5s<sup>2</sup>5p<sup>4</sup></small>

**Figura 3. 6.** Tabla periódica

Las [ecuaciones 3.14](#) y [3.16](#) se obtienen bajo la hipótesis de que  $E_F$  está lo suficientemente lejos de  $E_C$  y  $E_V$  ( $E_C - E_F > 3kT$  y  $E_F - E_V > 3kT$ ) en tal caso, el semiconductor se dice no degenerado y el producto de electrones y la concentración del hueco no depende de la energía de Fermi. En un semiconductor no dopado (intrínseco), la concentración del electrón y el hueco es igual y este valor es la concentración del portador intrínseco  $n_i$ .

$$n_i = \sqrt{n_0 p_0} = \sqrt{N_C N_V} e^{(E_V - E_C)/2kT} = \sqrt{N_C N_V} e^{-(E_g/2kT)} \quad (3.18)$$

Esta última ecuación representa la ley de la *acción de masas*: independientemente del dopaje, el producto de las densidades de huecos y electrones es una propiedad del material que no depende de la temperatura. Se debe tener en cuenta que cuanto más ancha es la banda prohibida, menor es la concentración del operador intrínseco.

Finalmente, el nivel de Fermi para un semiconductor intrínseco puede evaluarse igualando los electrones y la concentración del hueco dados por las [ecuaciones 3.14](#) y [3.16](#), respectivamente.

$$E_{F_i} = \frac{(E_C + E_V)}{2} + \frac{kT}{2} \ln \left( \frac{N_V}{N_C} \right) \quad (3.19)$$

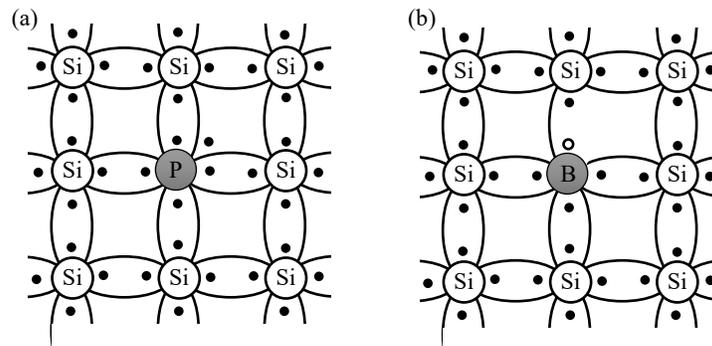
El nivel de Fermi está muy cerca del medio de la brecha de banda.

La concentración de portador intrínseco para Silicio ( $n_i \approx 10^{10} \text{ cm}^{-3}$ ) lo convierte en un aislante, para explotarlo como semiconductor es necesaria la introducción de impurezas o dopantes específicos.

### 3.1.6 Dopaje

La conductividad de un semiconductor puede variar introduciendo dopantes específicos. Con referencia a la parte de la tabla periódica esbozada en la [Figura 3. 6](#), se puede observar que el fósforo tiene cinco electrones de valencia ( $3s^2 3p^3$ ) mientras que el boro tiene tres electrones de valencia ( $3s^2 3p^1$ ).

Si se introducen átomos de fósforo en un cristal de silicio, uno de sus cinco electrones de valencia queda disponible para la conducción, los cuatro electrones restantes se unen con enlaces de covalencia de la red de silicio (véase la [Figura 3. 7a](#)). Este tipo de dopante se le dice donante. Del mismo modo, al introducir boro, sus tres electrones de valencia se unen con enlaces de covalencia de la red de silicio y queda un orificio (véase la [Figura 3. 7b](#)). Este tipo de dopante se les dice receptor.



**Figura 3. 7.** a Un dopaje de tipo n con Fosforo As, b dopaje tipo - p con Boro

Desde el punto de vista de los niveles de energía, la presencia de donantes introduce niveles de energía adicionales cerca de la banda de conducción (en unos pocos  $kT$ ), por lo tanto, la energía térmica puede permitir que el electrón agregado se mueva a la banda de conducción. Del mismo modo, la presencia de un receptor introduce niveles de energía adicionales cerca de la banda de valencia.

En caso de introducción del donante, los electrones son la fuente primaria de conducción y se dice que el semiconductor es de tipo-n, por el contrario, si se introduce un receptor, la conducción se debe al hueco, y se dice que el semiconductor es de tipo p.

Los átomos de los donantes ( $N_D$ ) o receptores ( $N_A$ ) suelen estar completamente ionizados, como consecuencia del semiconductor de tipo n  $n_0 \approx N_D$  y para el semiconductor tipo-p  $p_0 \approx N_A$ .

La presencia de dopante cambia el nivel de Fermi en comparación con un semiconductor intrínseco, este valor se puede volver a calcular mediante el uso de la [ecuación 3.19](#); para un semiconductor de tipo-n:

$$E_{F_n} = E_{F_i} + kT \ln \left( \frac{N_D}{n_i} \right) \quad (3.20)$$

Comparado con un semiconductor intrínseco el nivel de Fermi se incrementa. Para un semiconductor *tipo-p*.

$$E_{F_p} = E_{F_i} + kT \ln \left( \frac{N_A}{n_i} \right) \quad (3.21)$$

Y el nivel de Fermi es menor comparado con el semiconductor intrínseco. La ecuación (3.18) es válida incluso para semiconductores dopados, para un *tipo-n*:

$$n_i^2 = n_0 \cdot p_0 = p_0 \cdot N_D \quad (3.22)$$

Y la concentración de donadores puede ser expresada versus el nivel de Fermi para un semiconductor intrínseco:

$$n_0 = N_C e^{-(E_C - E_{F_n})/kT} = N_C e^{-(E_C - E_{F_i})/kT} e^{(E_{F_n} - E_{F_i})/kT} = n_i e^{(E_{F_n} - E_{F_i})/kT} \quad (3.23)$$

Dónde  $n_i$  se obtiene utilizando las [ecuaciones 3.18](#) y [3.19](#) con la posición  $N_C \approx N_V$ .

Para un *tipo-p*:

$$n_i^2 = n_0 \cdot p_0 = n_0 \cdot N_A \quad (3.24)$$

$$p_0 = n_i e^{-(E_{F_p} - E_{F_i})/kT} \quad (3.25)$$

En un semiconductor de *tipo n*, los electrones representan portadores mayoritarios y portadores minoritarios de huecos. Por lo general, si es necesario, su símbolo de concentración incluye un pedex para indicar el tipo de semiconductor. Por lo tanto, en un semiconductor de *tipo n* hay portadores mayoritarios  $n_n$  y portadores minoritarios  $p_n$ . En un semiconductor *tipo p*, hay portadores mayoritarios de  $p_p$  y portadores minoritarios  $n_p$ . Si es necesario, para especificar las condiciones de equilibrio se puede agregar un pedex adicional "o".

Cuando se realiza dopaje doble con donantes y receptores, el tipo de semiconductor se determina por la mayor concentración de impurezas.

Suponiendo la ionización total tanto de donantes como de receptores, si ocurre  $N_D > N_A$  y *tipo n* con electrones  $n_n \approx N_D - N_A$  y la concentración de huecos  $p_n$ , por el contrario, si  $N_A > N_D$  resulta un *tipo p* con  $p_p \approx N_A - N_D$  huecos y  $n_p$  concentración de electrones. De hecho, considerando, por ejemplo, el caso de  $N_D > N_A$ , siendo el semiconductor intrínseco y ambos dopantes neutros después del dopaje, la carga positiva final (debida a huecos y donantes ionizados) debe ser igual a la carga negativa (debido a electrones y donantes ionizados):

$$n_n + N_A = p_n + N_D \quad (3.26)$$

Utilizando la ley de acción de masas escrita en la forma  $n_i^2 = n_n p_n$  es posible obtener:

$$\begin{cases} n_n = \frac{1}{2} \left[ N_D - N_A + \sqrt{(N_D - N_A)^2 + 4n_i^2} \right] \\ p_n = \frac{n_i^2}{n_n} \end{cases} \quad (3.27)$$

por lo general la diferencia entre los donantes y receptores es mayor que la concentración intrínseca, que significa  $|N_D - N_A| > n_i^2$ .

### 3.1.7 Transporte del portador.

Hasta ahora, el Semiconductor ha sido considerado en el estado de equilibrio con  $n_0$  de electrones y  $p_0$  concentración de huecos. En esta condición, los electrones se mueven dentro de la celosía de

forma continua en todas las direcciones, pero, si se observa por un tiempo suficiente, su desplazamiento es nulo. La distancia media entre dos colisiones se conoce como el camino libre medio ( $L_c$ ) y el intervalo de tiempo correspondiente entre dos colisiones como el tiempo libre medio ( $\tau_c$ ). Esta situación es bosquejada esquemáticamente en la **Figura 3.8a**.

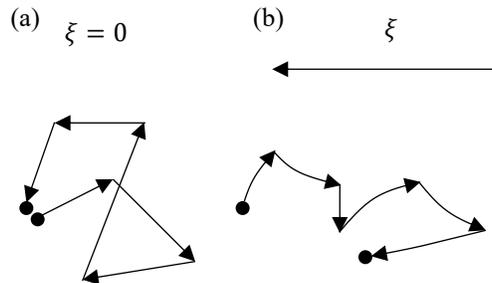
Mediante el uso de la masa efectiva  $m_n^*$ , la energía cinética de los electrones se puede expresar mediante la asignación de  $kt/2$  para cada grado de libertad.

$$\frac{1}{2} m_n^* v_{th}^2 = \frac{2}{3} kT \tag{3.28}$$

donde  $v_{th}$  es la velocidad libre media, a la temperatura ambiente ( $T = 300$  K) es igual a  $10^7$  cm/s tanto para Si como para GaAs. Considerando un valor típico de  $10^{-5}$  cm para la ruta de la media libre, sigue  $\tau_c = L_c/v_{th} = 10^{-12}$  s = 1 ps.

Si se aplica un campo eléctrico al semiconductor, se produce el desplazamiento de un conductor. El movimiento se representa esquemáticamente en la **Figura 3.8b**: el electrón (huevo) está continuamente disperso, pero hay un desplazamiento medio no nulo.

El movimiento de electrones y huecos se puede cambiar aplicando un campo eléctrico o variando localmente su concentración. En el primer caso, el proceso se denomina deriva, en esta última difusión.



**Figura 3.8.** a Movimiento térmico aleatorio, b movimiento térmico bajo campo eléctrico

### 3.1.7.1 Corriente de deriva.

Con respecto a la corriente de deriva, si se aplica un campo eléctrico a un semiconductor dopado uniformemente, la energía de las bandas no permanece constante, sino que se eleva en la dirección del campo. Los electrones se mueven en la dirección opuesta del campo hacia la energía más baja de la banda de conducción y los huecos hacia la energía más alta de la banda de valencia. El movimiento se representa esquemáticamente en la **Figura 3. 9**: el electrón (huevo) está continuamente disperso, pero hay un desplazamiento medio no nulo.

Los electrones tienen una velocidad de deriva en la dirección opuesta al campo eléctrico.

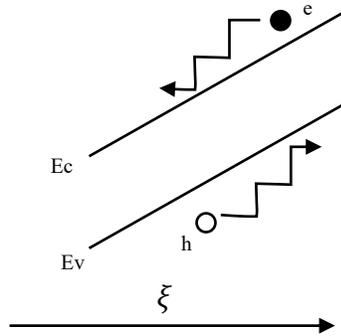
$$\vec{v}_{drift_n} = -\mu_n \vec{\xi} = \mu_n \nabla \phi \tag{3.29}$$

donde  $\phi$  es el potencial electrostático y  $\mu_n$  es el coeficiente de movilidad de electrones. Como consecuencia, se produce una corriente de deriva; su densidad es igual al producto de la velocidad de deriva para la densidad de electrones (o huecos) y para la carga de la unidad.

$$\vec{J}_{drift_n} = -qn\vec{v}_{drift_n} = qn\mu_n\vec{\xi} = -qn\mu_n\nabla\phi \quad (3.30)$$

Para un semiconductor cuya longitud (correspondiente a la dirección  $x$ ) es más grande que las otras dos dimensiones y el campo eléctrico está en la dirección  $x$ . Se puede usar una representación unidimensional (1D), como se muestra en la **Figura 3. 10**; [ecuación 3.30](#) se convierte en:

$$J_{drift_n} = -qnv_{drift_n} = qn\mu_n\xi = -qn\mu_n\frac{\partial\phi}{\partial x} \quad (3.31)$$



**Figura 3. 9.** Movimiento de niveles de energía y electrones (agujeros) en presencia de un campo eléctrico

donde  $\nabla\phi$  reduce a la derivada potencial a lo largo de  $x$ . La velocidad de deriva es proporcional al campo eléctrico  $\xi$  o a la derivada a lo largo de la dirección  $x$  del potencial electrostático  $\phi$ .

$$v_{drift_n} = -\mu_n\xi = \frac{\partial\phi}{\partial x} \quad (3.32)$$

De manera similar para los huecos:

$$\vec{J}_{drift_p} = -qp\vec{v}_{drift_p} = qp\mu_p\vec{\xi} = -qp\mu_p\nabla\phi \quad (3.33)$$

$$\vec{v}_{drift_p} = -\mu_p\vec{\xi} = \mu_p\nabla\phi \quad (3.34)$$

Los coeficientes que vinculan la velocidad al campo eléctrico se pueden evaluar igualando el impulso de la fuerza ( $-q\xi$ ) multiplicado por el intervalo de tiempo con la cantidad de movimiento lograda en el mismo tiempo. Para electrones:

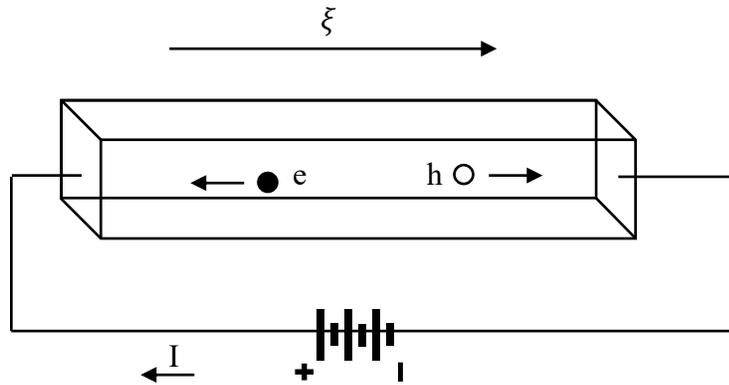
$$-q\xi\tau_n = m_n^*v_{drift_n} \quad (3.35)$$

$$\mu_n = \frac{q\tau_n}{m_n^*} \tag{3.36}$$

Y para los huecos:

$$-q\xi\tau_p = m_p^*v_{drift_p} \tag{3.37}$$

$$\mu_p = \frac{q\tau_p}{m_p^*} \tag{3.38}$$



**Figura 3. 10.** Semiconductor con campo eléctrico aplicado: se produce una corriente de deriva formada por agujeros y electrones

El coeficiente de movilidad tiene en cuenta el mecanismo de dispersión para la interacción reticular y la impureza ionizada.

### 3.1.7.2 Corriente de difusión

Cuando la concentración no es uniforme, los portadores tienden a pasar de regiones con alta concentración a regiones con baja concentración. Se produce una corriente de difusión y su densidad se da respectivamente para electrones y huecos por:

$$\vec{J}_{diff_n} = qD_n\nabla n \tag{3.39}$$

$$\vec{J}_{diff_p} = qD_p\nabla p \tag{3.40}$$

donde  $D_n$  y  $D_p$  son los coeficientes de difusividad para electrones y huecos respectivamente, y  $q$  es la carga de electrones.

Considerando un semiconductor largo donde la inyección se realiza en una superficie lateral, la corriente ocurre en dirección  $x$ :

$$J_{diff_n} = qD_n \frac{\partial n}{\partial x} \tag{3.41}$$

$$J_{diff_p} = -qD_p \frac{\partial p}{\partial x} \quad (3.42)$$

La corriente total, inducida en un semiconductor es la suma de la difusión y la corriente de deriva de electrones y huecos:

$$\vec{J}_n = \vec{J}_{drift_n} + \vec{J}_{diff_n} \quad (3.43)$$

$$\vec{J}_p = \vec{J}_{drift_p} + \vec{J}_{diff_p} \quad (3.44)$$

En equilibrio térmico, la corriente de difusión y la corriente de deriva deben balancearse. Comparando estos dos términos, se obtiene la relación de Einstein.

$$D_n = \left(\frac{kT}{q}\right) \mu_n \quad (3.45a)$$

$$D_p = \left(\frac{kT}{q}\right) \mu_p \quad (3.45b)$$

Mediante el uso de [ecuaciones 3.28](#), [3.36](#) y [3.38](#), el camino libre medio se obtiene frente a la difusividad y el tiempo libre medio:

$$L_n = \sqrt{D_n \tau_n} \quad (3.46a)$$

$$D_p = \left(\frac{kT}{q}\right) \mu_p \quad (3.46b)$$

Finalmente, se debe agregar un término adicional debido al campo de desplazamiento dieléctrico.

$$\vec{J}_{disp} = \frac{\partial \vec{D}}{\partial t} = \varepsilon \frac{\partial \vec{\xi}}{\partial t} \quad (3.47)$$

donde  $\varepsilon$  es la permitividad eléctrica del semiconductor. Sin embargo, se debe considerar que en el caso de una célula fotovoltaica que se opera en CD, este término puede despreciarse.

Para silicio a  $T = 300$  K y concentración de impurezas de  $10^{16} \text{ cm}^{-3}$  la movilidad de electrones es de aproximadamente  $1400 \text{ cm}^2/\text{Vs}$  y, en consecuencia, la difusividad es de aproximadamente  $35 \text{ cm}^2/\text{s}$ .

Los valores más bajos se muestran con orificios cuya movilidad es de aproximadamente  $490 \frac{\text{cm}^2}{\text{Vs}}$  y, por consiguiente, la difusividad es de aproximadamente  $10 \text{ cm}^2/\text{s}$ . Otro semiconductor como GaAs tiene valores mayores de alrededor de cinco veces. La movilidad y la difusividad disminuyen con la concentración de impurezas.

### 3.1.7.3 Resistividad del semiconductor

Considerando las [ecuaciones 3.30](#) y [3.33](#), la densidad de corriente total debida a un campo eléctrico viene dada por:

$$J = J_n + J_p = (qn\mu_n + qp\mu_p)\xi \quad (3.48)$$

Esta ecuación define la resistividad del semiconductor como:

$$\rho = \frac{1}{\sigma} = \frac{1}{(qn\mu_n + qp\mu_p)} \quad (3.49)$$

### 3.1.8 Ecuaciones fundamentales del semiconductor

El semiconductor obedece a dos leyes fundamentales: la ecuación de Poisson y la ecuación de continuidad.

La ecuación de Poisson proviene de la ley de Gauss para la electricidad, la cual establece que, en un volumen de control, la divergencia del campo de desplazamiento eléctrico  $\vec{D}$  es igual a la densidad de carga libre  $\rho_s$ . Para un medio lineal, isotrópico y homogéneo  $\vec{D} = \epsilon\vec{E}$ .

La ecuación de continuidad es una ley de conservación local. Afirma que una cantidad (no importa qué: partículas, cargas o personas) en un volumen dado puede cambiar en el tiempo solo para una variación de flujo de la cantidad a través de la superficie que contiene el volumen.

#### 3.1.8.1 Ecuación de Poisson

La forma general de un semiconductor fue propuesta por Van Roosbroeck en 1950:

$$\nabla \cdot \epsilon\vec{\xi} = \rho_s \quad (3.50)$$

donde  $\rho_s$  es la densidad de carga espacial. Para un semiconductor,  $\rho_s$  se puede expresar como la contribución de la carga positiva (densidad de agujeros más los donantes ionizados) y la carga negativa (densidad de electrones y aceptores ionizados).

$$\nabla \cdot \epsilon\vec{\xi} = q(p + N_D - n - N_A) \quad (3.51)$$

Esta última ecuación se puede particularizar a lo largo de la dirección  $x$ :

$$\epsilon \frac{d\xi}{dx} = q(p + N_D - n - N_A) \quad (3.52)$$

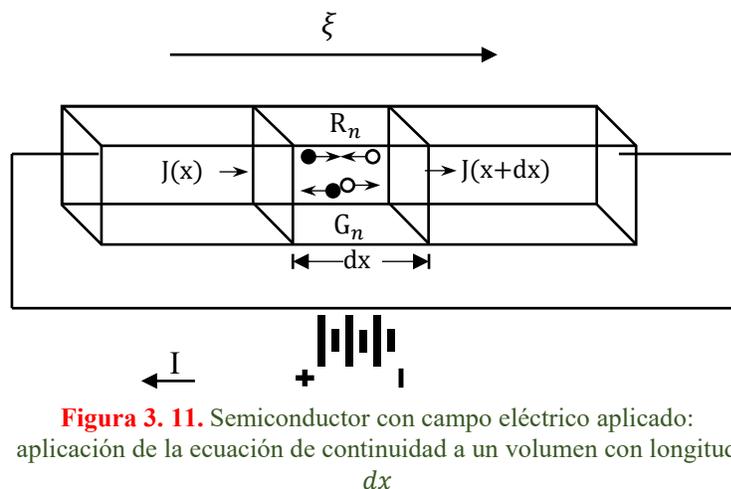
### 3.1.8.2 Ecuación de la continuidad

La ecuación de continuidad puede escribirse tanto para electrones como para agujeros. La forma general de los electrones es:

$$\nabla \cdot \vec{J}_n = q \left( R_n - G_n + \frac{\partial n}{\partial t} \right) \quad (3.53)$$

El primer miembro es la divergencia de la densidad de corriente de electrones. Representa el flujo de electrones a través de un volumen arbitrario.

El segundo miembro es la variación del número de electrones durante el tiempo. Se compone de un primer término que da la velocidad de la recombinación  $R_n$ . El proceso de recombinación ocurre cuando un electrón cae desde la banda de conducción a la banda de valencia y desaparece un par electrón – hueco.



**Figura 3. 11.** Semiconductor con campo eléctrico aplicado: aplicación de la ecuación de continuidad a un volumen con longitud  $dx$

Varios mecanismos de recombinación son posibles: por defectos en la banda prohibida, recombinación radiante y un término adicional denominado recombinación de Auger.

El primer tipo de recombinación ocurre cuando un electrón o un hueco realiza una transición a un nivel de energía parásita dentro de la banda prohibida y se emiten los fonones. En la recombinación radiante, una transición de la banda de conducción a la banda de valencia origina una emisión de fotones.

En la recombinación de Auger, un electrón y un hueco se recombinan en una transición de banda a banda, pero en este caso la energía resultante se transfiere a otro electrón u orificio, en este último caso están involucradas tres partículas.

El segundo término de la [ecuación 3.53](#),  $G_n$ , da la velocidad de generación, esto ocurre para absorción de luz o inyección de portador.

El tercer término es la variación de tiempo de la densidad de electrones. En la representación 1D, como se representa en la **Figura 3. 11**, la ecuación de continuidad para los electrones es:

$$\frac{\partial J_n}{\partial x} = q \left( R_n - G_n + \frac{\partial n}{\partial t} \right) \quad (3.54)$$

Para huecos:

$$\nabla \cdot \vec{J}_p = q \left( G_p - R_p - \frac{\partial p}{\partial t} \right) \quad (3.55)$$

$$\frac{\partial J_p}{\partial x} = q \left( G_p - R_p + \frac{\partial p}{\partial t} \right) \quad (3.56)$$

### 3.1.9 Ecuaciones de difusión de portadores minoritarios

Una representación 1D del semiconductor se considera a continuación. Para condiciones de estado estable, la [ecuación 3.54](#) se puede usar despreciando la derivada de tiempo. Por otra parte, la derivada de densidad de corriente considerando las [ecuaciones 3.30](#) y [3.39](#) viene dado por:

$$\frac{d}{dx} J_n = q(R_n - G_n) = \frac{d}{dx} \left( -qn\mu_n\xi + qD_n \frac{\partial n}{\partial x} \right) \quad (3.57)$$

y para huecos

$$\frac{d}{dx} J_p = q(R_p - G_p) = \frac{d}{dx} \left( -qn\mu_p\xi + qD_p \frac{\partial n}{\partial x} \right) \quad (3.58)$$

Para el campo eléctrico pequeño, la contribución de la derivación del espacio del campo eléctrico puede despreciarse. Esto equivale a despreciar la corriente de deriva en comparación con la corriente de difusión. Bajo inyección de bajo nivel, la tasa de recombinación para electrones se puede escribir como la diferencia entre los portadores minoritarios en esa posición y su valor de equilibrio dividido por la vida útil del portador minoritario.

$$R_n = \frac{p_n - p_{n0}}{\tau_p} = \frac{\Delta p_n}{\tau_p} \quad (3.59)$$

Debe observarse que la recombinación de electrones depende de los portadores minoritarios correspondientes. Como cuestión de hecho, la recombinación puede ocurrir cuando un portador minoritario está disponible. Para un material de tipo  $n$ , de la [ecuación 3.58](#) se obtiene la ecuación de difusión del portador minoritario correspondiente:

$$D_p \frac{d^2}{dx^2} \Delta p_n - \frac{\Delta p_n}{\tau_p} = -G(x) \quad (3.60)$$

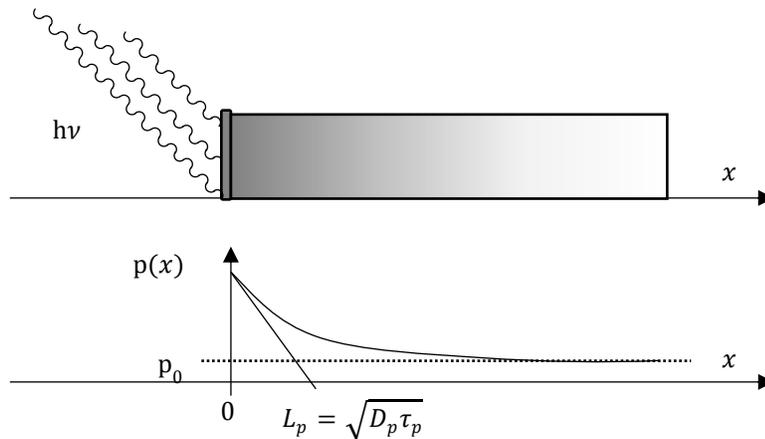
Para un material de *tipo p*, la ecuación de difusión del portador minoritario correspondiente es:

$$D_n \frac{d^2}{dx^2} \Delta n_p - \frac{\Delta n_p}{\tau_n} = -G(x) \tag{3.61}$$

Cuando se enciende un semiconductor n dopado en una superficie lateral, se realiza una inyección de portador, la concentración del portador minoritario viene dada por [ecuación 3.57](#) imponiendo  $G(x) = 0$  con condiciones de contorno:

$$\begin{cases} p_n(x = 0) = p_n(0) \\ p_n(x \Rightarrow \infty) = p_{n0} \end{cases} \tag{3.62}$$

En la superficie, la inyección causa un valor constante de la concentración de portadores minoritarios  $p_n(0)$ , por el contrario, para  $x$  lo suficientemente lejos de la superficie de inyección, la concentración vuelve al valor de equilibrio  $p_{n0}$ .



**Figura 3. 12.** Representación de una barra semiconductor con una luz incidente en la superficie lateral que provoca la inyección de carga: la concentración de carga se relaja con la constante espacial  $L_p$ , para  $x$  lo suficientemente lejos de la zona de inyección se restablece la concentración de equilibrio

La solución es:

$$p_n(x) = p_{n0} + [p_n(0) - p_{n0}]e^{-x/L_p} \tag{3.63}$$

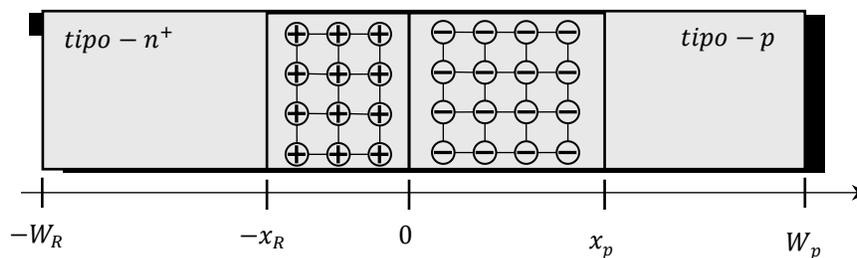
La concentración de operadores minoritarios se relaja con la ley exponencial. La constante del espacio de relajación viene dada por las [ecuaciones 3.46a](#) y [3.46b](#). La corriente de difusión resultante se puede obtener mediante [ecuación 3.41](#).

Debe observarse que, si la longitud del semiconductor es aproximadamente  $L_p$ , debe imponerse una condición de contorno diferente en la superficie correspondiente donde los portadores de concentración minoritaria son diferentes del valor de equilibrio.

La **Figura 3.12** muestra una situación típica en la cual la luz incidente en la superficie lateral causa inyección de carga, su concentración se relaja con constante espacial  $L_p$ , si la distancia es suficientemente lejos de la zona de inyección, la concentración de equilibrio se restablece y la concentración de carga es igual al equilibrio valor  $p_0$ .

### 3.1.10 Unión P-N

Una unión  $p - n$  puede concebirse conceptualmente como un semiconductor dopado de *tipo n* y *tipo p* que tienen una superficie en común. Cuando ambos semiconductores están separados, son eléctricamente neutros. Tan pronto como entran en contacto, los portadores mayoritarios de semiconductores de *tipo n* (los electrones) comienzan a difundirse en el semiconductor de *tipo p* y viceversa. Como resultado, cerca de la superficie de separación entre los dos semiconductores, en el semiconductor de *tipo n*, los huecos provenientes del semiconductor de *tipo p* tienden a combinarse con los electrones y la carga positiva de los correspondientes donantes ionizados no es más compensada por los portadores mayoritarios. Dentro de la región de *tipo n*, cerca de la unión, donde no hay más cargas mayoritarias, se observa un agotamiento y la zona correspondiente permanece con cargas positivas fijas.



**Figura 3.13.** Representación esquemática de una unión p-n

Del mismo modo, en el lado del *tipo p*, los electrones procedentes del semiconductor *tipo n* tienden a combinarse con los orificios y la carga negativa de los receptores ionizados no es compensada más, por los portadores mayoritarios. Dentro de la región de *tipo p*, cerca de la unión, donde no hay más cargas de la mayoría, se observa un agotamiento y la zona correspondiente permanece con cargas negativas fijas.

A medida que se descubren las cargas fijas, se produce un campo eléctrico y el proceso de difusión se ralentiza. Se dibuja una unión  $p - n$  en la **Figura 3.13** en representación 1D; el origen ( $x = 0$ ) es la superficie de unión,  $x_p$  y  $W_p$  son el límite de agotamiento al final de la región de *tipo p*, mientras que  $-x_n$  y  $-W_n$  son el límite de agotamiento al final de la región de *tipo n*. Debe observarse que, si un semiconductor está más dopado que el otro (generalmente indicado con ápice +), la mayor cantidad de portador libre difundido en el otro semiconductor causa un agotamiento más prolongado.

Se supone un dopaje uniformado y no degenerado y los dopantes están completamente ionizados. Toda la zona en la que hay una carga no compensada fija se denomina región de agotamiento o región de carga espacial. Las zonas restantes se pueden considerar como neutrales (a menudo denominadas cuasineutral). El campo eléctrico debido a las cargas fijas origina una diferencia de

potencial electrostático llamado voltaje incorporado. La ecuación de Poisson [3.50](#) se puede reescribir como:

$$\frac{d\xi}{dx} = -\nabla^2\phi = \frac{q}{\varepsilon}(p_0 + N_D - n_0 - N_A) \quad (3.64)$$

donde  $\phi$  es el potencial electrostático,  $p_0$  y  $n_0$  son la concentración de equilibrio del hueco y del electrón,  $N_D$  es la concentración de donantes ionizados (cargas fijas positivas), y  $N_A$  es la concentración de receptores ionizados (cargas fijas negativas).

La [ecuación 2.64](#) puede particularizarse para cada zona. En la zona cuasineutral, la carga neta es nula y sigue a  $\nabla^2\phi = 0$ , dentro de la región de agotamiento cargada positiva  $\nabla^2\phi = -qN_D/\varepsilon$  y dentro de la región de agotamiento cargada negativamente  $\nabla^2\phi = qN_A/\varepsilon$ .

El voltaje incorporado es la diferencia de potencial entre los puntos  $x = -x_n$  y  $x = x_p$ .

$$V_{bi} = \phi(-x_n) - \phi(x_p) \quad (3.65)$$

Suponiendo que  $\phi(x_p) = 0$ . El proceso del potencial a lo largo de  $x$  es:

$$\left\{ \begin{array}{ll} V_{bi} & \text{for } x \leq -x_n \\ V_{bi} - \frac{qN_D}{2\varepsilon}(x + x_n)^2 & \text{for } -x_n < x < 0 \\ \frac{qN_A}{2\varepsilon}(x - x_p)^2 & \text{for } 0 \leq x < x_p \\ 0 & \text{for } x \geq x_p \end{array} \right. \quad (3.66)$$

Aceptando que la segunda y la tercera relación en [ecuación 3.66](#) dan los mismos valores de  $\phi$  para  $x = 0$  y la continuidad del campo eléctrico en el mismo punto, que da:

$$x_n N_D = x_p N_A \quad (3.67)$$

la longitud de la zona de agotamiento se obtiene frente a la tensión incorporada.

$$W_D = x_n + x_p = \sqrt{\frac{2\varepsilon}{q} \left( \frac{N_A + N_D}{N_A N_D} \right) V_{bi}} \quad (3.68)$$

La [ecuación 3.68](#) dice que la longitud de la zona de agotamiento depende del voltaje incorporado. Para el funcionamiento del diodo, este valor puede disminuirse aplicando una tensión positiva en el lado  $p$  (polarización directa) o ampliándose mediante una tensión positiva en el lado  $n$  (polarización inversa).

En particular, en caso de polarización directa, la tensión incorporada  $V_{bi}$  se reduce en una cantidad igual a la tensión aplicada y el campo eléctrico correspondiente cambia las condiciones de

equilibrio permitiendo que los huecos pertenecientes a la zona  $p$  pasen a través de la unión y alcancen la zona  $n$  y los electrones que pertenecen a una zona, para pasar a través de la unión y alcanzar la zona  $p$ , también.

La longitud de la zona de agotamiento se puede calcular mediante [ecuación 3.68](#) usando el nuevo valor de  $(V_{bi} - V)$  donde  $V$  es la tensión aplicada.

Para inyección de carga de bajo nivel, ocurre una corriente de difusión debido a cargas minoritarias como se describe en la [ecuación 3.63](#) para agujeros en la zona  $n$ . La corriente correspondiente se calcula por medio de [\(3.42\)](#).

$$J_{diff_p} = qD_p \frac{\partial p}{\partial x} = q \frac{D_p}{L_p} [p_n(0) - p_{n0}] e^{-x/L_p} \quad (3.69)$$

Ahora, teniendo en cuenta que la corriente total permanece igual a lo largo de  $x$  porque dentro de la región cuasineutral no hay una tasa apreciable de generación-recombinación, es fácil evaluar la [ecuación. 3.69](#) para  $x = 0$ .

$$J_{diff_p}(0) = q \frac{D_p}{L_p} [p_n(0) - p_{n0}] = q \frac{D_p}{L_p} \nabla p_n \quad (3.70)$$

Cabe señalar que, debido a que la corriente es constante a lo largo de la dirección  $x$ , disminuyendo la corriente de los portadores minoritarios, se produce una corriente complementaria debida a los portadores mayoritarios a lo largo de  $x$ . Para  $x \gg L_p$ , la corriente solo se debe a portadores mayoritarios.

$$qp\mu_p \vec{\xi} = qD_p \frac{dp}{dx} \quad (3.71)$$

Al usar la relación de Einstein ([3.45a](#)) y asumiendo  $V_T = kT/q$ , [ecuación 3.71](#) se convierte en:

$$\xi = \frac{V_T}{p} \frac{dp}{dx} = - \frac{dV}{dx} \quad (3.72)$$

En cuanto a la concentración de huecos, se debe observar que, dentro de la zona  $p$ , se asume el valor correspondiente al equilibrio térmico  $p_{p0}$ , en el borde de la zona  $n$ , cerca de la zona de agotamiento es igual a  $p_n(0)$ . Reescribiendo la [ecuación 3.72](#) e integrando sobre la zona de agotamiento:

$$\int_{p_{p0}}^{p_n(0)} \frac{dp}{p} = - \frac{1}{V_T} \int_0^{V_{bi}-V} dV \quad (3.73)$$

Se deduce que:

$$p_n(0) = p_{p0} e^{-\frac{(V_{bi}-V)}{V_T}} \quad (3.74)$$

Finalmente, la relación entre  $p_{p0}$  y  $p_{n0}$  se puede encontrar integrando la [ecuación 3.72](#) bajo condición de no sesgo (significa  $V = 0$ ) obteniendo:

$$p_{n0} = p_{p0} e^{-(V_{bi})/V_T} \quad (3.75)$$

A partir del cual

$$p_n(0) - p_{n0} = p_{n0} \left[ e^{\frac{V}{V_T}} - 1 \right] \quad (3.76)$$

y la [ecuación 3.70](#) se convierte en:

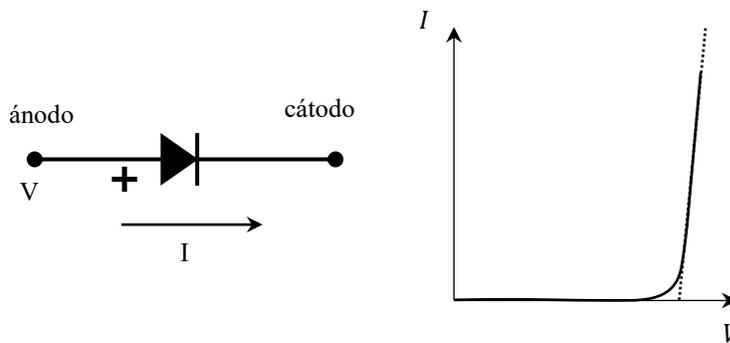
$$J_{diff_p}(0) = q \frac{D_p p_{n0}}{L_p} \left[ e^{\frac{V}{V_T}} - 1 \right] \quad (3.77)$$

Finalmente, la corriente de unión se determina sumando la contribución de electrones y multiplicando para el área de sección A.

$$I = I_s \left( e^{\frac{V}{V_T}} - 1 \right) \quad (3.78)$$

Dónde

$$I_s = Aq \left( \frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n} \right) \quad (3.79)$$



**Figura 3.14.** Símbolo eléctrico y voltaje en función del diodo de corriente

La [ecuación 3.78](#) se conoce como ecuación del diodo de Shockley. Cuando la tensión de polarización directa se acerca a  $V_{bi}$ , la zona de agotamiento tiende a desaparecer y la corriente está limitada también por el semiconductor y el contacto óhmico. En este caso, la característica de voltaje versus corriente se aproxima por línea recta.

Cuando se aplica un sesgo inverso, significa que se aplica un voltaje positivo la zona de contacto  $n$ , y la [ecuación 3.78](#) puede ser utilizado todavía. Como cuestión de hecho, el término exponencial es insignificante y una corriente de saturación inversa dada por [ecuación 3.79](#) se obtiene. En este caso, la pequeña corriente obtenida solo está dada por portadores generados dentro de la unión y no depende del sesgo inverso aplicado. La **Figura 3. 14** muestra la tensión frente a la característica de corriente de un diodo.

De acuerdo con lo explicado dada anteriormente, está claro que un diodo permite que la corriente pase de una zona  $p$  a una zona  $n$  cuando está polarizada hacia adelante. El contacto óhmico que pertenece a la zona  $p$  se llama ánodo, mientras que el contacto óhmico que pertenece a la zona  $n$  se llama cátodo.

### 3.1.11 Capacitancia de la unión P-N

En [sección 3.1.10](#), se ha explicado que la corriente cerca de la unión se debe a la inyección del portador minoritario. Cuando la tensión de polarización cambia, la concentración de carga correspondiente también varía, pero de forma similar a un condensador, esta concentración requiere un intervalo de tiempo finito para alcanzar un nuevo equilibrio.

La capacitancia equivalente puede calcularse considerando la difusión del portador minoritario en [ecuación 3.63](#) en el que la inyección se debe a la polarización de tensión directa de unión para la que se inyectan electrones en la región  $p$ . Si la región  $p$  está más dopada que la región  $n$ , se puede suponer que la corriente se da solo por agujeros inyectados en la región  $n$ .

La variación de carga total en comparación con la condición de equilibrio se puede calcular como:

$$\begin{aligned} Q &= A \int_0^{\infty} q[p_n(x) - p_{n0}]dx = A \int_0^{\infty} q[p_n(0) - p_{n0}]e^{-x/L_p}dx = \\ &= Aq[p_n(0) - p_{n0}] = AqL_p\Delta p_n \end{aligned} \quad (3.80)$$

donde  $A$  es el área transversal de la unión. Al comparar la [ecuación 3.80](#) con [3.70](#) y usando la [ecuación 3.46b](#), se deduce que:

$$J_{diff_p} = \frac{Q}{\tau_p} \quad (3.81)$$

Significa que la corriente es proporcional al exceso de cargas minoritarios. Para la polarización directa, la capacitancia de la unión se denomina capacitancia de difusión; para la polarización de la unión inversa, la capacitancia se denomina capacitancia de transición y tiene un valor más bajo que la capacitancia de difusión; además, bajo operación de células PV, la polarización inversa de

la unión no ocurre, por el contrario, la evaluación de capacitancia de difusión proporciona información sobre el comportamiento dinámico.

Al usar la [ecuación 3.81](#), bajo condiciones estáticas:

$$C_D = \frac{dQ}{dV} = \tau_p \frac{dI}{dV} = \tau_p g \quad (3.82)$$

donde  $g$  es la conductancia diferencial de la unión; se puede obtener sobre la base de la [ecuación 3.78](#) despreciando los términos correspondientes a la corriente de saturación inversa.

$$g = \frac{dI}{dV} = \frac{d}{dV} \left[ I_s \left( e^{\frac{V}{V_T}} - 1 \right) \right] = \frac{I_s}{V_T} e^{V/V_T} = \frac{I}{V_T} \quad (3.83)$$

Entonces:

$$C_D = \tau_p \frac{I}{V_T} \quad (3.84)$$

La capacidad de difusión es proporcional a la corriente. Si los huecos y los electrones contribuyen a la corriente, la capacidad de difusión es la suma de dos términos debido a los electrones y los agujeros, respectivamente. Finalmente, comparando las [ecuaciones 3.82](#) y [3.84](#), se puede ver que la constante de tiempo del circuito equivalente  $RC$  es igual a la vida útil de las cargas minoritarias:

$$\frac{C_D}{g} = \tau_p \quad (3.85)$$

Si el voltaje aplicado varía con el tiempo, la capacidad de difusión dinámica se define sobre la base de la variación de carga en el intervalo de tiempo  $dt$  como:

$$i = \frac{dQ'}{dt} = \frac{dQ'}{dv} \cdot \frac{dv}{dt} = C_D' \frac{dv}{dt} \quad (3.86)$$

Durante el intervalo de tiempo  $dt$ , solo la carga cerca de la unión varía, de hecho, la difusión del portador requiere más tiempo para alcanzar una nueva condición de equilibrio.

Por esta razón, la variación de carga es menor en comparación con la condición estática y se espera un menor valor de capacidad de difusión. En cualquier caso, el valor dinámico de la capacitancia de difusión depende de la forma de onda del voltaje aplicado; se puede obtener una expresión de esta tensión para una caja sinusoidal con amplitud  $V_m$  y pulsación  $\omega$  superpuesta a una polarización fija  $V_1$ .

$$v(t) = V_1 + V_m e^{j\omega t} \quad (3.87)$$

si  $V_m$  es una señal pequeña, se puede despreciar el efecto no lineal de la curva  $I - V$  y la corriente resultante tiene la forma

$$i(t) = I_1 + gV_m e^{j\omega t} + j\omega C_D' V_m e^{j\omega t} \quad (3.88)$$

El exceso de concentración de portadores minoritarios contiene, además del término debido a la polarización estática, un término dependiente del tiempo adicional:

$$\Delta p_n(x, t) = \Delta p_n(x) e^{j\omega t} \quad (3.89)$$

Utilizando las [ecuaciones 3.57](#) y [3.58](#) donde no hay generación de carga y se desatiende la derivada del campo eléctrico, se obtiene la ecuación diferencial para las cargas minoritarias en exceso:

$$D_p \frac{d^2}{dx^2} \Delta p_n - \frac{\Delta p_n}{\tau} - \frac{d(\Delta p_n)}{dt} = 0 \quad (3.90)$$

Sustituyendo la [ecuación 3.89](#) en la [ecuación 3.90](#) y teniendo en cuenta que de la [ecuación 3.46a](#)  $L_p = \sqrt{D_p \tau_p}$ :

$$\frac{d^2}{dx^2} \Delta p_n = \frac{1 + j\omega \tau_p}{L_p^2} \Delta p_n \quad (3.91)$$

Lo que resulta en:

$$\Delta p_n = k e^{\frac{(1+j\omega \tau_p)^{\frac{1}{2}} x}{L_p}} e^{j\omega t} \quad (3.92)$$

la concentración resultante de exceso de portadores minoritarios está dada por el término de tiempo constante definido por la [ecuación 3.63](#) más el término variable en el tiempo definido por la [ecuación 3.92](#).

$$\Delta p_n = p_{n0} \left( e^{\frac{V_1}{V_T}} - 1 \right) e^{-x/L_p} + k e^{\frac{(1+j\omega \tau_p)^{\frac{1}{2}} x}{L_p}} e^{j\omega t} \quad (3.93)$$

donde  $k$  es una constante que se puede encontrar al igualar la ley de empalme [ecuación 3.76](#) considerando que la tensión de unión está definida por la [ecuación 3.87](#).

$$\Delta p_n = p_{n0} \left( e^{\frac{V_1 + V_m e^{j\omega t}}{V_T}} - 1 \right) e^{-x/L_p} \quad (3.94)$$

con la expresión de concentración de portadores minoritarios en exceso resultante [ecuación 3.93](#) ambos evaluados para  $x = 0$ .

Ya que, para  $y \ll 1$ ,  $e^y \approx 1 + y$ , suponiendo  $(V_m/V_T)e^{j\omega t} = y$ , se obtiene la siguiente expresión de  $k$ :

$$k = p_{n0} \frac{V_m}{V_T} e^{\frac{V_1}{V_T}} \quad (3.95)$$

Finalmente, la corriente de unión se obtiene usando la [ecuación 3.42](#) para  $x = 0$ .

$$I_p(0) = \frac{AqD_p p_{n0}}{L_p} \left( e^{\frac{V_1}{V_T}} - 1 \right) + \frac{AqD_p p_{n0} V_m e^{V_1/V_T}}{V_T} \cdot \frac{\sqrt{1 + j\omega\tau_p}}{L_p} e^{j\omega t} \quad (3.96)$$

Debe notarse que el primer término del segundo miembro de la [ecuación 3.96](#) corresponde a la corriente estática de la unión  $p - n$ ,  $I_1$ .

Para bajas frecuencias, corresponde a  $\omega\tau_p \ll 1$ , la aproximación  $\sqrt{1 + j\omega\tau_p} \approx 1 + j\omega\tau_p/2$  es válido, entonces la [ecuación 3.96](#) da:

$$I_p(0) = I_1 + \frac{AqD_p p_{n0} V_m e^{V_1/V_T}}{L_p} \left( 1 + \frac{j\omega\tau_p}{2} \right) e^{j\omega t} \quad (3.97)$$

$$\left\{ \begin{array}{l} g = g_0 \frac{AqD_p p_{n0} V_m e^{V_1/V_T}}{L_p} \\ C_D = \frac{g\tau_p}{2} \end{array} \right. \quad (3.98)$$

Debe notarse que la conductancia es la misma del valor estático  $g_0$  pero la capacitancia se divide entre dos. Ambos dependen del voltaje de polarización estática mediante el término exponencial.

Para altas frecuencias correspondientes a  $\omega\tau_p \gg 1$

$$\sqrt{1 + k\omega\tau_p} = \sqrt{j\omega\tau_p} = \sqrt{(\omega\tau_p)} e^{j\frac{\pi}{4}} = \sqrt{\frac{\omega\tau_p}{2}} (1 + j)$$

Y la [ecuación 3.96](#) da:

$$\left\{ \begin{array}{l} g = g_0 \sqrt{\frac{\omega\tau_p}{2}} \\ C_D = g_0 \sqrt{\frac{\tau_p}{2\omega}} \end{array} \right. \quad (3.99)$$

En este último caso, tanto la capacitancia difusa como la conductancia dependen de la frecuencia.

### 3.1.12 La celda Fotovoltaica

La célula fotovoltaica (PV) es básicamente una unión  $p - n$  con una región de depleción central y dos zonas cuasineutrales laterales. Al final de cada zona se coloca un contacto eléctrico. La zona más fuertemente dopada se llama zona emisora y la otra es la zona base. Esta última región también se denomina región absorbente porque la gran parte de la luz incidente se absorbe aquí. A diferencia de un diodo, la célula fotovoltaica está diseñada para permitir que se generen pares de electrodos y huecos dentro de la unión debido a la luz incidente.

El objetivo de esta sección es definir la ley que vincula el voltaje y la corriente de una célula fotovoltaica, incluida la dependencia de la luz incidente. En una palabra, se debe encontrar la corriente total debida tanto a los electrones como a los agujeros.

$$I = A[J_p(x) + J_n(x)] \quad (3.100)$$

donde  $A$  es la superficie del semiconductor. Con este objetivo, la ecuación de continuidad de electrones (2.54) en estado de equilibrio puede integrarse sobre la región de agotamiento obteniendo:

$$\int_{-x_N}^{x_p} \frac{\partial J_n}{\partial x} dx = J_n(x_p) - J_n(-x_N) = q \int_{-x_N}^{x_p} [R_n(x) - G_n(x)] dx \quad (3.101)$$

Ahora, evaluando la [ecuación 3.100](#) para  $x = -x_N$  y utilizando la [ecuación 3.101](#) para deducir  $J_n(-x_N)$ , se obtiene la siguiente ecuación:

$$I = A \left[ \int_{-x_N}^{x_p} J_p(-x_N) + J_n(x_p) + q \int_{-x_N}^{x_p} R_n(x) dx \right] \quad (3.102)$$

Esta última ecuación contiene cuatro términos. La primera es la corriente de densidad de huecos evaluada al final de la zona de agotamiento en la región de *tipo n*, la segunda es la corriente de densidad de electrones evaluada al final de la zona de agotamiento en la región de *tipo p*. Estos dos términos se refieren a operadores minoritarios. El tercer término es la tasa de generación óptica y el último término representa la tasa de recombinación en la región de carga espacial.

#### 3.1.12.1 Densidad de corriente de los portadores minoritarios

Las ecuaciones de difusión de portadores minoritarios en las regiones de *tipo n* cuasineutral y de *tipo p* tienen que resolverse con condiciones de frontera adecuadas. En particular, para el semiconductor de *tipo n* [ecuación 3.60](#) con condición de frontera en  $x = -W_N$  y  $x = -x_N$ ; y para el semiconductor *tipo p* [ecuación 3.61](#) con condición de frontera en  $x = W_P$  y  $x = x_P$ . Se

ha explicado en la [sección 3.1.8](#) que el dopaje cambia el nivel de energía de Fermi. Por lo tanto, la ley de acción masiva se puede escribir como:

$$n_i^2 = p n e^{-(E_{F_n} - E_{F_p})/kT} = p n e^{-(\Delta E_{F_{np}})/kT} \quad (3.103)$$

donde  $\Delta E_{F_{np}}$  es la diferencia entre el nivel de Fermi debido a  $n$  el dopaje y el nivel de Fermi debido al dopaje.

Suponiendo, que la mayoría de los portadores mantengan una concentración constante en sus respectivas zonas cuasineutrales, para el *tipo n* significa que el nivel de energía de Fermi es constante para  $-W_n \leq x \leq -x_n$  y para *tipo p* que el nivel de energía de Fermi es constante para  $x_p \leq x \leq W_p$ , por lo tanto, se puede suponer en todas partes dentro del semiconductor:

$$\Delta E_{F_{np}} = E_{F_n}(-W_n) - E_{F_p}(-W_p) = E_{F_n}(x) - E_{F_p}(x) = qV \quad (3.104)$$

Ahora, estando en la región cuasineutral semiconductor de *tipo n*  $n = N_D$ , se obtiene la primera condición límite:

$$p_n(-x_n) = \frac{n_i^2}{N_D} e^{qV/kT} \quad (3.105)$$

Y, estando en la región cuasineutral semiconductor de *tipo p*  $p = N_A$ , se tiene que:

$$n_p(x_p) = \frac{n_i^2}{N_A} e^{qV/kT} \quad (3.106)$$

Cabe señalar la presencia del factor  $e^{qV/kT}$  en las [ecuaciones 3.105](#) y [3.106](#), que estarán presentes en la expresión final de la corriente generada por la celda FV. Las dos últimas condiciones de contorno se obtienen teniendo en cuenta que el contacto frontal en la zona *tipo n* está construido por una rejilla metálica con una superficie frontal efectiva  $S_{F,eff}$ , y la recombinación efectiva está dada por:

$$\frac{d}{dx} \Delta p = \frac{S_{F,eff}}{D_p} \Delta p(-W_n) \quad (3.107)$$

Esta situación es diferente de un contacto óhmico ideal en el que  $\Delta p = 0$ . Para el semiconductor de *tipo p*, el contacto posterior es un contacto bastante óhmico, sin embargo, si se introduce una región más delgada y dopada cerca de este contacto para aumentar el número de portadores minoritarios recogidos, una condición similar a la ecuación 3.107 puede imponerse introduciendo la superficie posterior efectiva  $S_{BSF}$ :

$$\frac{d}{dx} \Delta n = \frac{S_{BSF}}{D_n} \Delta p(W_p) \quad (3.108)$$

### 3.1.12.2 Tasa de generación óptica

Como se explicó anteriormente, solo los fotones con longitud de onda  $\lambda \leq hc/E_g$  pueden contribuir a generar pares electrón – hueco. La tasa de generación depende de un factor de sombreado  $s$  de la cuadrícula, de la reflectancia  $r(\lambda)$ , del coeficiente de absorción  $\alpha(\lambda)$  y del flujo de fotones incidente  $f(\lambda)$  de acuerdo con la [ecuación 3.109](#).

$$\int_{-x_N}^{x_p} G_n(x) dx = (1 - s) \int_{\lambda} [1 - r(\lambda)] f(\lambda) (e^{-\alpha(W_N - x_N)} - e^{-\alpha(W_N - x_p)}) d\lambda \quad (3.109)$$

### 3.1.13 Tasa de recombinación

Este término depende de la longitud de la zona de agotamiento y de la concentración del portador intrínseco, es inversamente proporcional a la vida útil efectiva en la región de agotamiento  $\tau_D$ . Además, está presente un término adicional en el que la exponencial contiene la tensión multiplicada por  $(q/2kT)$  es igual a aproximadamente 19.23 J/C para  $T = 300$  K.

$$\int_{-x_N}^{x_p} R_n(x) dx = \frac{W_D n_i}{\tau_D} (e^{(qV/2kT)} - 1) \quad (3.110)$$

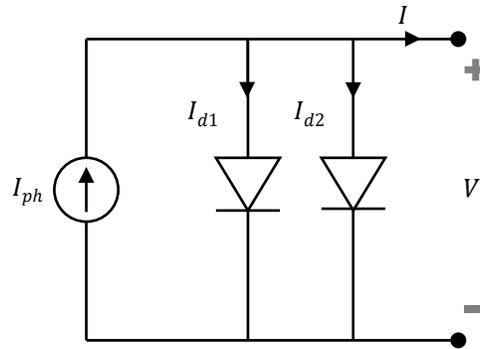
Al resolver la ecuación de difusión del portador minoritario con las condiciones de contorno y considerando la generación y recombinación, la ley actual contra voltaje de la célula fotovoltaica se obtiene después de cálculos no triviales, no incluidos aquí.

$$I = I_{ph} - I_{s1} (e^{qV/kT} - 1) - I_{s2} (e^{qV/2kT} - 1) \quad (3.111)$$

El primer término es la corriente de cortocircuito y recoge las contribuciones de las dos zonas cuasineutrales y de la zona de agotamiento.

El segundo término contiene la corriente de saturación oscura debido a la recombinación en la región cuasineutral  $I_{s1}$ . Debe observarse que en la [ecuación 3.111](#) este fenómeno se considera ideal. Cuando se tiene en cuenta la no idealidad de la difusión, se debe introducir un factor de idealidad en el exponencial.

Finalmente, el último término contiene la corriente de saturación oscura debido a la recombinación en la región de carga espacial  $I_{s2}$ . Este último término depende del sesgo porque es proporcional a la longitud de la zona de agotamiento.



**Figura 3. 15.** Modelo de circuito físico de una célula fotovoltaica

### 3.1.14 Modelo físico de la celda FV

La [ecuación 3.111](#) reproduce los fenómenos físicos de una célula fotovoltaica. Puede ser considerado como una ley de corriente de Kirchhoff (KCL) escrita para un nodo en un circuito eléctrico para el cual la suma de las corrientes que fluyen hacia ese nodo es igual a la suma de corrientes que fluyen de ese nodo. Sobre la base de esta ley, un circuito equivalente puede deducirse. Representa un modelo de circuito físico de una célula PV. Este circuito se dibuja en la [Figura 3. 15](#).

Cabe señalar que la corriente de salida es la suma de una corriente dada por un generador eso depende de la irradiación solar menos la corriente que fluye a través de los dos diodos.

La primera corriente corresponde a  $I_{ph}$  en la [ecuación 3.111](#), la segunda corriente corresponde a  $I_{s1}(e^{qV/kT} - 1) = I_{d1}$  y la tercera corriente corresponde a  $I_{s2}(e^{qV/2kT} - 1) = I_{d2}$ .

Como cuestión de hecho, el segundo y el tercer término de la [ecuación 3.111](#) puede ser consideradas como ecuaciones de diodo de Shockley. Finalmente, la tensión de salida se obtiene por el sesgo directo de los diodos debido al generador de corriente.

Durante las condiciones de operación, cuando ocurre la radiación solar, la corriente del generador fluye a través de los diodos y aparece una tensión en los terminales. Si no se aplica carga este voltaje es un voltaje de "circuito abierto", es decir, el voltaje de un voltaje directamente de la unión  $p - n$  polarizada y es el valor máximo alcanzable por una célula PV. Si una carga está conectada, una parte de la corriente del generador fluye hacia la carga, el voltaje disminuye y se suministra energía eléctrica a la carga. El proceso de conversión es terminado.

A partir de la radiación solar, se obtuvo energía eléctrica.

Cabe señalar que si la carga se eleva (corresponde a una resistencia más baja) la corriente también aumenta y el voltaje disminuye; la potencia suministrada alcanza un máximo y luego disminuye hasta el estado de cortocircuito.

Cuando no hay radiación solar presente, la corriente generada es nula y, por consiguiente, la tensión en los terminales. Sin embargo, esto no corresponde a un comportamiento de cortocircuito, por el contrario, la celda PV no permite el flujo de corriente negativa impuesto por los circuitos externos.

### 3.1.15 Tipos de semiconductores

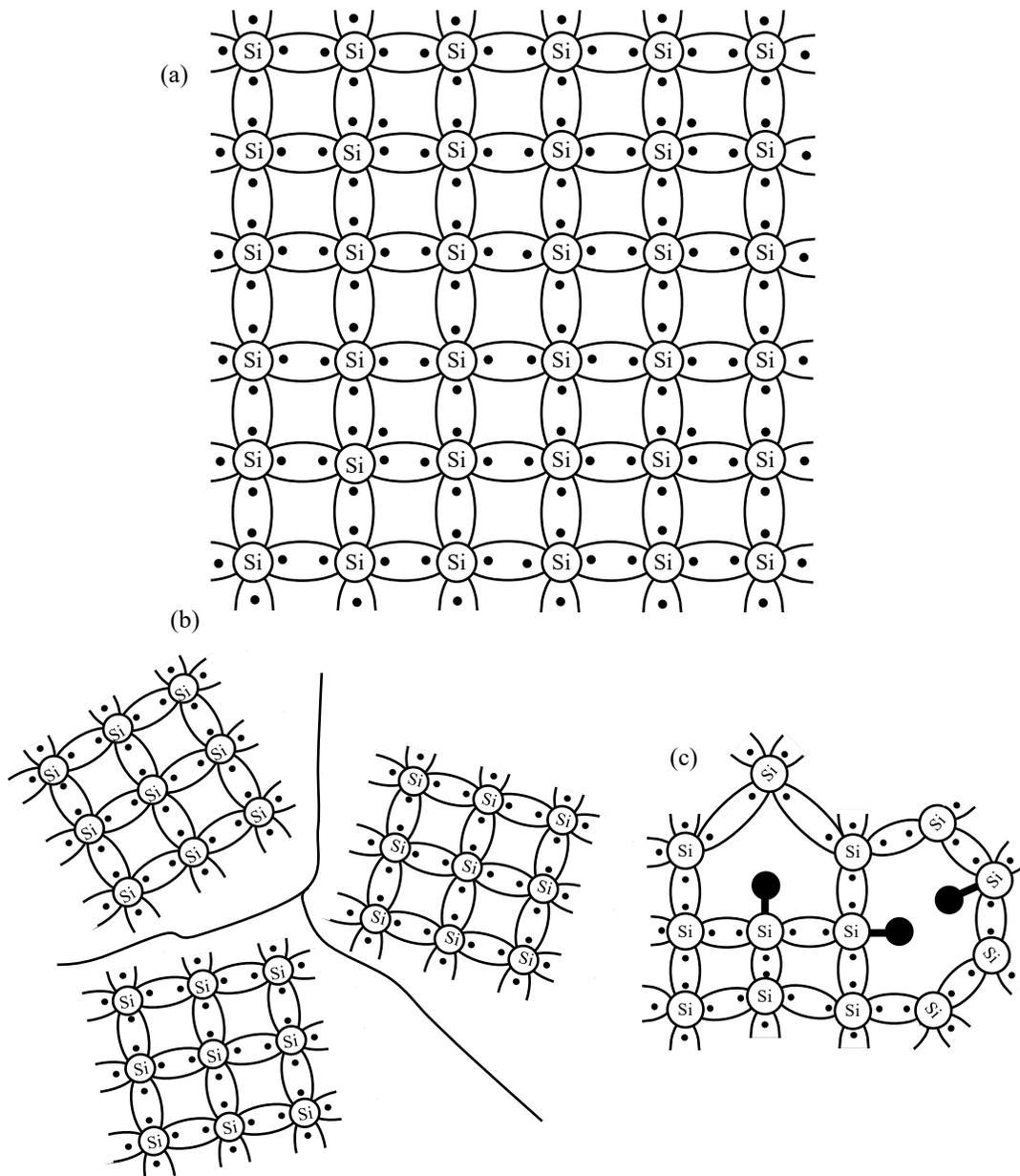
El silicio se puede organizar en diferentes formas para obtener una célula fotovoltaica dependiendo de la estructura del cristal. Siguiendo la clasificación en base al tamaño de grano plano, se pueden distinguir los microcristalinos, policristalinos y multicristalinos. En particular, el material microcristalino exhibe un grano menor de 1  $\mu\text{m}$ , policristalino menor de 1  $\text{mm}$  y multicristalino menor de 10  $\text{cm}$ .

En la figura 2.16 se dibuja una representación esquemática de la estructura del silicio cristalino, policristalino y amorfo.

#### 3.1.15.1 Silicio cristalino

El silicio cristalino se considera como una estructura ideal donde el patrón es regular en toda la superficie. Toda la teoría explicada anteriormente se desarrolla con referencia a esta estructura. La principal ventaja consiste en la mayor potencia solar producida por irradiación solar. Con el silicio monocristalino, se espera una eficiencia de conversión de energía que oscile entre el 20 y el 24%, con GaAs, se espera una eficiencia de conversión de energía del 20 al 29%.

El silicio cristalino, por otro lado, es caro debido al proceso de fabricación. Por esta razón, se han desarrollado varias estructuras alternativas de silicio más baratas.



**Figura 3.16.** Estructura de silicio cristalino (a), policristalino (b) y amorfo (c)

### 3.1.15.2 Multicristalino

El silicio policristalino y multicristalino se puede producir mediante una técnica menos sofisticada que la cristalina. Sin embargo, en este caso, se debe tener en cuenta la presencia de límites de grano. En particular, el rendimiento de la celda se reduce porque en los límites que las portadoras fluyen está bloqueado, la estructura de nivel está alterada y la corriente que fluiría a través de la unión p-n se desvía.

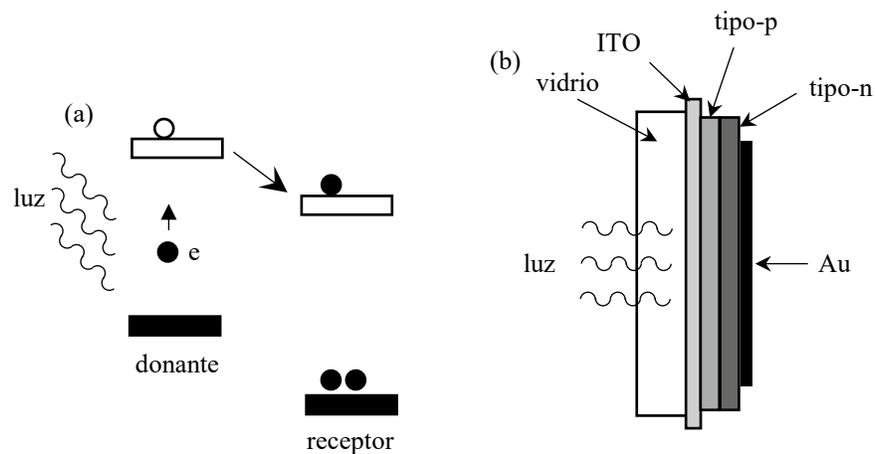
Se han ideado algunos remedios como, por ejemplo, el uso de granos de pocos milímetros para cubrir toda la distancia desde la parte posterior hasta la parte delantera de la celda con un número mínimo de granos. Con silicio policristalino, se espera una eficiencia de conversión de energía que va del 13 al 18%.

### 3.1.15.3 Amorfo

El silicio amorfo presenta una estructura menos regular con enlaces insatisfechos. Estos enlaces "suspendidos" son pasivados por hidrógeno al permitir el dopaje (de lo contrario es imposible) y elevar el intervalo de banda de 1.1 eV de silicio cristalino a 1.7 eV; de esta forma, se pueden absorber fotones de mayor energía y el espesor requerido del material es menor. Como consecuencia, el silicio amorfo se puede usar como una forma de "película delgada" depositada sobre vidrio u otros sustratos para aplicaciones de bajo costo.

La estructura de banda de los materiales amorfos es similar al material cristalino a corta distancia y se puede definir un espacio de movilidad, en el que se produce la conducción.

Sin embargo, hay una gran cantidad de estados de energía localizada dentro de la brecha de movilidad, que corresponden a las colas de la banda y enlaces de desgarre que hacen diferente el comportamiento amorfo en comparación con el silicio cristalino.



**Figura 3. 17.** Transferencia de electrones de material donante a receptor (a), sección de una célula orgánica (b)

#### 3.1.15.4 Película delgada

Las llamadas tecnologías de película delgada, donde se depositan células de película delgada sobre un sustrato extraño, están indicadas para una transición desde otras tecnologías para la fabricación de módulos.

Es posible, al aumentar la relación al silano de hidrógeno en el gas del que se deposita el silicio amorfo, obtener un material microcristalino en el que las columnas de cristalitas están separadas por una región amorfa.

En realidad, las tecnologías de película delgada establecidas utilizan silicio amorfo y película de Si nanocristalina (a-Si), telururo de cadmio (CdTe) y diseleniuro de cobre y indio (CIG), tienen una eficacia del 8-13% para a-Si, de 10-17% para CdTe, y para 10-19% para CIG. Además, se espera un costo reducido. De hecho, en un módulo de Si cristalino, aproximadamente el 40% del costo se debe a la fabricación de obleas o cintas de células solares.

#### 3.1.15.5 Célula solar polimérica

Además del material inorgánico anteriormente citado como silicio (Si), arseniuro de galio (GaAs), telururo de cadmio (CdTe) y cadmio-indio-seleniuro (CIS), los esfuerzos para reducir los costos debido al silicio de grado solar y los procesos de fabricación han llevado a la explotación de materiales orgánicos de película delgada.

En particular, los polímeros y moléculas conjugados son fáciles de producir, tienen flexibilidad mecánica como los plásticos y pueden exhibir propiedades electrónicas similares a los conductores y los semiconductores. Además, los avances recientes en técnicas de litografía blanda como la impresión por chorro de tinta y la impresión por microcontacto han aumentado el interés hacia la fabricación de dispositivos integrados sobre sustratos rígidos y flexibles que tienen un área grande. Sin embargo, las propiedades de los semiconductores orgánicos e inorgánicos son significativamente diferentes. En un semiconductor orgánico, como se explicó anteriormente, la absorción de un fotón genera una pareja agujero-electrón y se transportan por su gran movilidad y campo eléctrico interno.

Por el contrario, en la disociación de semiconductores orgánicos en cargas libres, los portadores no se producen a temperatura ambiente. Como consecuencia, se necesitan dos materiales diferentes y se crean cargas como un efecto de la transferencia de electrones fotoinducidos de uno a otro material.

En particular, después de la absorción de la irradiación solar, para la cual es deseable una correspondencia entre el espectro solar y el espectro de absorción, se requiere un material donante (D) para dar el electrón excitado al material receptor (A). Como resultado, se produce un catión radical del donante ( $D^+$ ) y un anión radical del aceptor ( $A^-$ ).

Además, los procesos competitivos como la fluorescencia y el deterioro no radiativo deben inhibirse y la transferencia de cargas fotogeneradas debe ser ayudada. Por estas razones, el electrón transferido debe ralentizarse tanto como sea posible.

Estas últimas consideraciones están esquematizadas en la figura **Figura 3. 17a**. Aquí, un electrón se excita por absorción de luz al orbital molecular ocupado más alto (HOMO) en el material donante y luego se transfiere al orbital molecular desocupado más bajo (LUMO) en el material

aceptor. Ahora, un electrón adicional está presente en el material aceptor y aparece un orificio en el material donante. La **Figura 3. 17b** dibuja la celda fotovoltaica. La luz pasa a través de un vidrio y un electrodo transparente (ITO); los electrodos son diferentes pero ambos son metálicos para recolectar cargas y transferirlas a circuitos externos.

Entre las células orgánicas, las más exitosas parecen ser: células sensibilizadas con colorante, células de doble capa y células de heterounión a granel.

Las células sensibilizadas con colorante utilizan un colorante orgánico adsorbido en la superficie de un semiconductor inorgánico de banda ancha para la absorción de luz y la inyección de fotoelectrones en la banda de conducción del semiconductor. Las celdas de doble capa consisten en un sandwich de una sola capa de colorante orgánico entre dos electrodos diferentes.

Finalmente, en las células de heterounión a granel, los materiales de tipo p y n se mezclan entre sí, pero, debido a la tendencia intrínseca de los materiales de polímero a permanecer separados en la dimensión nanométrica, se crean uniones a lo largo del material.

# Capítulo

## 4

### Estadística para datos experimentales

#### Definiciones.

**Datos** son las observaciones recolectadas (como mediciones, géneros, respuestas de encuestas).

**Estadística** es una colección de métodos para planear experimentos, obtener datos, y después organizar, resumir, presentar, analizar, interpretar y llegar a conclusiones basadas en los datos.

**Población** es la colección completa de todos los elementos (puntuaciones, personas, mediciones, etcétera) a estudiar. Se dice que la colección es completa, pues incluye a todos los sujetos que se estudiarán.

**Censo** es la colección de datos de cada uno de los miembros de la población.

**Muestra** es un subconjunto de miembros seleccionados de una población.

**Estadístico** es una medición numérica que describe algunas características de una muestra.

**Datos cuantitativos** consisten en números que representan conteos o mediciones.

**Datos cualitativos** (o categóricos o de atributo) se dividen en diferentes categorías que se distinguen por alguna característica no numérica.

**Datos discretos** resulta cuando el número de posibles valores es un número finito, o bien, un número que puede contarse. (Es decir, el número de posibles valores es 0, 1, 2, etcétera).

**Datos continuos** (numéricos) resultan de un infinito de posibles valores que pueden asociarse a puntos de alguna escala continua, cubriendo un rango de valores sin huecos ni interrupciones.

#### 4.1 Diseño de experimentos.

Es necesario entender que el método usado para reunir los datos es extremadamente importante, y debemos reconocer que la *aleatoriedad* resulta importante en particular [59].

- Si los datos muestrales no se reúnen de manera adecuada, éstos podrían resultar inútiles por completo, de tal forma que ninguna cantidad de tortura estadística los salvaría [59].
- La *aleatoriedad* por lo general juega un papel crucial para determinar cuáles son los datos a reunir [59].

Los métodos estadísticos se rigen por los datos. Por lo regular obtenemos datos de dos fuentes distintas: los estudios observacionales y los experimentos [59].

### Definiciones.

En un **estudio observacional**, observamos y medimos características específicas, aunque no intentamos manipular a los sujetos que estamos estudiando.

En un **experimento** aplicamos algún tratamiento y luego procedemos a observar sus efectos sobre los sujetos.

En un **estudio tradicional**, los datos se observan, miden y reúnen en un solo momento.

En un **estudio retrospectivo** (o de **control de caso**), los datos se toman del pasado (a través del examen de registros, entrevistas y otros medios).

En un **estudio prospectivo** (o **longitudinal o cohorte**), los datos se reúnen en el futuro y se toman de grupos (llamados cohortes) que comparten factores comunes.

Existe una distinción importante entre el muestreo realizado en estudios retrospectivos y estudios prospectivos. En los estudios retrospectivos regresamos en el tiempo a reunir datos a cerca de características resultantes que nos conciernen, como un grupo de conductores que murieron en accidentes automovilísticos y otro grupo de conductores que no murieron en este tipo de accidentes. En los estudios prospectivos vamos adelante en el tiempo siguiendo grupos con un factor causal potencial y grupo que no lo tienen, como un grupo de conductores que utilizan teléfonos celulares y un grupo de conductores que no usan teléfonos celulares. Las tres definiciones se aplican a los estudios observacionales, aunque por ahora nos enfocamos en los experimentos. Los resultados de los experimentos algunas veces se empobrecen a causa de la confusión [59].

### Definiciones.

La **confusión** ocurre en un experimento cuando el experimentador no es capaz de distinguir entre los efectos de diferentes factores.

Utilice un tamaño de muestra que sea lo suficientemente grande para distinguir la verdadera naturaleza de cualquiera de los diferentes efectos, y obtenga la muestra usando un método adecuado, como uno basado en la aleatoriedad [59].

### Definiciones.

En una **muestra aleatoria** los miembros de una población se seleccionan de manera que cada *miembro individual* tiene la misma posibilidad de ser elegido.

Una **muestra aleatoria simple** del tamaño de  $n$  sujetos, se selecciona de manera que cada posible *muestra del mismo tamaño  $n$*  tenga la misma posibilidad de ser elegida.

En una **muestra aleatoria** los miembros de una población se seleccionan de manera que cada *miembro individual* tiene la misma posibilidad de ser elegido.

Una **muestra aleatoria simple** del tamaño de  $n$  sujetos, se selecciona de manera que cada posible *muestra del mismo tamaño  $n$*  tenga la misma posibilidad de ser elegida.

En el **muestreo sistemático**, elegimos algún punto de partida y luego seleccionamos cada  $k$ -ésimo (por ejemplo, cada quincuagésimo) elemento en la población.

Con el **muestreo de conveniencia**, simplemente se utilizan resultados que sean muy fáciles de obtener.

Con el **muestreo estratificado**, subdividimos la población en al menos dos diferentes subgrupos (o estratos) que comparten las mismas características (por ejemplo, el género o la categoría de edad) y después realizamos un muestreo de cada subgrupo (o estrato).

En el **muestreo por racimos**, primero dividimos el área de la población en secciones (o racimos), después seleccionamos aleatoriamente algunos de estos racimos, y luego elegimos a *todos* los miembros de los racimos seleccionados.

Un **error de muestreo** es la diferencia entre el resultado de una muestra y el verdadero resultado de la población; tal error es consecuencia de las posibles fluctuaciones de las muestras.

Un **error no de muestreo** ocurre cuando los datos de una muestra se obtienen, registran o analizan de forma incorrecta (como cuando se selecciona una muestra sesgada o predispuesta, cuando se usa un instrumento de medición defectuosa o cuando se cometen errores al copiar los datos).

## 4.2. Descripción, exploración y comparación de datos.

### 4.2.1. Características importantes de los datos.

1. **Centro:** valor representativo o promedio que indica la localización de la mitad del conjunto de los datos.
2. **Variación:** medida en que la cantidad de los valores de los datos varían entre sí.
3. **Distribución:** naturaleza o forma de la distribución de los datos (tales como normales, uniformes o sesgados).
4. **Datos distantes:** valores muestrales que están muy alejados de la vasta mayoría de los demás valores de la muestra.
5. **Tiempo:** características cambiantes de los datos a través del tiempo.

Los métodos de este capítulo suelen denominarse métodos de **estadística descriptiva**, porque su objetivo es resumir o *describir* las características importantes de un conjunto de datos. Más adelante, utilizaremos métodos de **estadística inferencial**; lo haremos cuando usemos datos muestrales para hacer inferencias (o generalizaciones) acerca de una población. Con la estadística inferencial realizamos una deducción que va más allá de los datos

conocidos. La materia de estadística tiene dos divisiones generales: la descriptiva y la inferencial; este capítulo trata los conceptos básicos de la estadística descriptiva [59].

#### 4.2.2. Distribuciones de frecuencias.

Cuando se trabaja con un conjunto grande de datos, con frecuencia es útil organizarlos y resumirlos por medio de la construcción de una tabla que liste los distintos valores posibles de los datos (ya sea de forma individual o por grupos), junto con las frecuencias correspondientes (o conteos) correspondientes [59].

#### Definiciones.

Los **límites de clase inferiores** son las cifras más pequeñas que pueden pertenecer a las diferentes clases.

Los **límites de clase superiores** son las cifras más grandes que pueden pertenecer a las diferentes clases.

Las **fronteras de clase** son las cifras utilizadas para separar las clases, aunque sin los espacios creados sin por los límites de clase. Se obtienen de la siguiente manera: se determina el tamaño del espacio entre el límite de clase superior de una clase y el límite de clase inferior de la siguiente.

Las **marcas de clase** son los puntos medios de las clases. Cada marca de clase se calcula sumando el límite de clase inferior con el límite de clase superior y dividiendo la suma entre dos.

La **anchura de clase** es la diferencia entre dos límites de clase inferiores consecutivos o dos fronteras de clase inferiores consecutivas.

**Tabla 4. 1.** Tabla de frecuencias de la irradiancia solar.

Irradiancia (W/m <sup>2</sup> )	Frecuencia
0-159	15
160-319	1
320-479	2
480-639	2
640-799	4

#### Procedimiento de construcción de una distribución de frecuencias

Las distribuciones de frecuencias se construyen por las siguientes razones: **1.** Es posible resumir conjuntos grandes de datos, **2.** Se logra cierta comprensión respecto de la naturaleza de los datos, y **3.** Se llega a tener un avance para construir graficas importantes (tales como

*histogramas*, que se presentarán en la siguiente sección). A continuación, se presenta el procedimiento básico para construir las distribuciones de frecuencia manualmente:

1. Decida el número de clases que desea tener. Debe ser entre 5 y 20, y deben utilizarse números enteros o redondos.
2. Calcule

$$\text{Anchura de clase} \approx \frac{(\text{valor más alto}) - (\text{valor más bajo})}{\text{números de clases}} \quad (4.1)$$

Redondee el resultado para obtener un número más adecuado (generalmente se redondea hacia *arriba*). Es probable que necesite cambiar el número de clases, pero la prioridad debe ser utilizar valores que sean fáciles de comprender [59][60][61].

3. Punto de partida: comience por elegir un número para el límite inferior de la primera clase. Elige el valor del dato más bajo o un valor conveniente que sea un poco más pequeño.
4. Con el uso del límite más bajo de la primera clase y anchura de clase, proceda a listar los demás límites de clase inferior. (Sume la anchura de clase a punto de partida para obtener el segundo límite de clase inferior para obtener el segundo límite de clase inferior. Después, sume la anchura de clase el segundo límite de clase inferior para obtener el tercero y así sucesivamente).
5. Anote los números inferiores de clase en la columna vertical y luego proceda a anotar los límites superiores de clase, que puedan identificarse con facilidad.
6. Ponga una marca en la clase apropiada para cada dato. Utilice las marcas para obtener la frecuencia total de cada clase.

#### 4.2.3. Visualización de los datos.

El principal objetivo de este capítulo es aprender técnicas importantes para investigar las características “CVDDT” importantes de los conjuntos de datos: centro, variación, distribución, datos distantes y cambios a lo largo del tiempo. En esta sección continuaremos el estudio de las distribuciones por medio de la introducción de gráficas, que son dibujos de distribuciones. El objetivo no es simplemente la construcción de gráficas, sino más bien aprender algo acerca de los conjuntos de datos, es decir, comprender la naturaleza de sus distribuciones [59].

### Histogramas

**Histograma** es una gráfica de barras en donde la escala horizontal representa escalas de valores de datos y la escala vertical representa frecuencias. Las alturas de las barras

corresponden a frecuencias, en tanto que las barras se dibujan de manera adyacente (sin espacios entre ellas) [59].

### Diagrama de dispersión

Un **diagrama de dispersión** es una gráfica de datos apareados  $(x, y)$ , con un eje  $x$  horizontal y un eje  $y$  vertical. Los datos se aparean de tal forma que cada valor de un conjunto de datos corresponde a un valor de un segundo conjunto de datos [59].

### Gráficas de series de tiempo

Los **datos de series de tiempo** son aquellos que se reúnen en diferentes momentos [59].

#### 4.2.4. Medidas de tendencia central

El principal objetivo de este capítulo es lograr manejar las herramientas básicas para medir y describir diferentes características de un conjunto de datos. Cuando describimos, exploramos y comparamos conjuntos de datos, las siguientes características suelen ser extremadamente importantes: centro, variación, distribución, datos distantes, cambios a través del tiempo [59].

#### Definiciones.

**Medida de tendencia central:** valor que se encuentra en el centro o a la mitad de un conjunto de datos.

Hay muchas formas distintas de determinar el centro; por lo tanto, tenemos diferentes definiciones de las medidas de tendencia central, incluyendo media, mediana, moda y mitad del rango [59].

### Media

La media (aritmética) generalmente es la más importante de todas las medidas numéricas utilizadas para describir datos; constituye lo que la mayoría de la gente denomina *promedio*.

#### Definiciones.

**Media aritmética** (de un conjunto de puntajes): medida de tendencia central que se obtiene sumando los puntajes y dividiendo el total entre el número de puntajes.

$$media = \bar{x} = \frac{\sum x}{n}$$

Para un conjunto de datos de una muestra se denota  $\bar{x}$  y  $\mu$  si se trata de una población.

### Notación

$\Sigma$  Denota la *sumatoria* de un conjunto de valores.

$x$  es la *variable* que suele utilizarse para representar los valores de datos individuales.

$n$  representa el *número de valores* de la muestra.

$N$  representa el *número de valores* de una población.

$\bar{x} = \frac{\Sigma x}{n}$  es la media de un conjunto de valores *muéstrales*.

$\mu = \frac{\Sigma x}{N}$  es la media de todos los valores de una *población*.

## Mediana

### Definición.

**Mediana** (de un conjunto de datos): medida de tendencia central que implica el valor que está en medio, cuando los valores originales de los datos se presentan en orden de magnitud creciente (o decreciente). La mediana suele denotarse con  $\tilde{x}$ .

## Moda

### Definición.

**Moda** (de un conjunto de datos que suele denotarse como  $M$ ): valor que ocurre con mayor frecuencia.

## Mitad del rango

### Definición.

**Mitad del rango:** medida de tendencia central que constituye el valor que está a medio camino, entre el puntaje más alto y el más bajo, en el conjunto original de datos. Se calcula sumando el valor máximo con el mínimo y luego dividiendo dicha suma entre 2, como en la siguiente fórmula.

$$\text{mitad del rango} = \frac{\text{Valor máximo} + \text{valor mínimo}}{2}$$

**Regla de redondeo**

Una regla sencilla para redondear respuestas es la siguiente:

**Aumentar una posición decimal más a las que están presentes en el conjunto original de datos.**

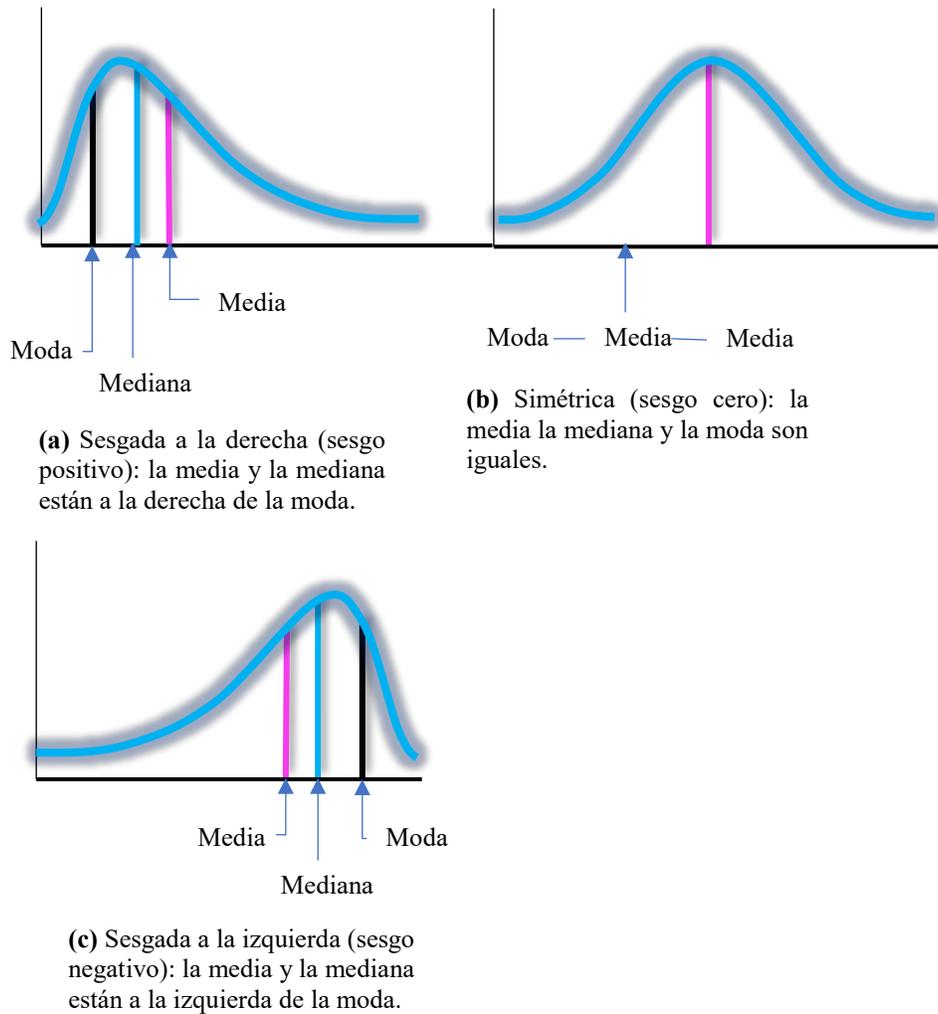
**Sesgo****Definición.**

Una distribución de datos está **sesgada** si no es simétrica y se extiende más hacia un lado que hacia el otro. (Una distribución de datos es **simétrica** si la mitad izquierda de su histograma es aproximadamente una imagen en espejo de su mitad derecha).

Los datos sesgados a la izquierda (que también se denomina como sesgo negativo) poseen una cola izquierda más larga, en tanto que la media y la mediana se encuentran a la izquierda de la moda. Aunque no siempre es posible predecirlo, los datos sesgados a la izquierda suelen tener una media menor a la mediana, como sucede en la **Figura 4. 1a**. Los datos sesgados a la derecha (lo que también se denomina sesgo positivo) poseen una cola derecha más larga, mientras que la media y la mediana se encuentran a la derecha de la moda. Nuevamente, aunque no siempre es posible predecirlo, en los datos sesgados a la derecha, la media suele estar a la derecha de la mediana, como en la **Figura 4. 1c**.  
Ejemplo 4.1. Medidas de tendencia central. En la **Tabla 4. 2**.

**Tabla 4. 2.** Medidas de tendencia central para datos de radiación solar durante las horas del día

Media (promedio)	$418.7 \text{ W/m}^2$
Mediana	$386 \text{ W/m}^2$
Mitad del rango	$603.5 \text{ W/m}^2$



**Figura 4. 1.** Sesgo

### 4.2.5. Medidas de variación

#### Rango

**Definición.**

El **rango** de un conjunto de datos es la diferencia entre el valor máximo y el valor mínimo.

$$Rango = (\text{valor máximo}) - (\text{valor mínimo})$$

#### Desviación estándar de una muestra

La desviación estándar es, por lo general, la medida de variación más importante y útil.

**Definición.**

**Desviación estándar** (de un conjunto de valores muestrales): medida de variación de los valores con respecto a la media. Es un tipo de desviación promedio de los valores, con respecto a la media, que se calcula utilizando las fórmulas 3.2 o 3.3.

Fórmula 4.2	$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$	Desviación estándar de la muestra
Fórmula 4.3	$s = \sqrt{\frac{n\sum(x^2) - \sum(x)^2}{n(n - 1)}}$	Fórmula abreviada de la desviación estándar de la muestra

**Desviación estándar de una población**

Para calcular la desviación estándar  $\sigma$  de una población, se utiliza una fórmula ligeramente diferente: en lugar de dividir entre  $n - 1$ , se hace entre el tamaño  $N$  de la población, como en la siguiente expresión:

$\sigma = \sqrt{\frac{\sum x - \mu}{N}}$	Desviación estándar de una población
--	--------------------------------------

**Varianza de una muestra y una población**

Usamos el término variación como una descripción general de la cantidad que varían los valores entre sí [59].

**Definición.**

**Varianza** (de un conjunto de valores): medida de variación igual al cuadrado de la desviación estándar.

Varianza muestral: cuadrado de la desviación estándar  $s$ .

Varianza poblacional: cuadrado de la desviación estándar poblacional  $\sigma$ .

Comparación de la variación en diferentes poblaciones.

**Definición.**

**Coefficiente de variación** o CV de un conjunto de datos muestrales o poblacionales, expresado como porcentaje, describe la desviación estándar relativa a la media, y está dada de la siguiente forma:

Muestra  

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Población  

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Ejemplo 4.2 Medidas de variación. En la **Tabla 4.3** se muestran medidas de variación para el mes de enero y febrero, como se puede apreciar en la tabla, el mes de enero presenta mayor variación o dispersión en los datos [60].

**Tabla 4.3.** Medidas de variación para muestras del mes de enero y febrero

		Mes	
		Enero	Febrero
Medida de variación	Rango	864	965
	Des. Estándar	247.3	311.3
	Varianza	61152.5	96891.4
	CV	76.68%	76.84%

**4.3. Tercer y cuarto momento.**

4.3.1. Tercer momento (sesgo).

Tercer momento (skewness), es una medida de la asimetría de la distribución, conocido también como el sesgo, el valor obtenido es adimensional y se define como:

$$S_k = \frac{\sum_{i=1}^n [(x_i - \bar{x})^3]^2}{[\sum_{i=1}^n (x_i - \bar{x})^2]^3} \tag{4.4}$$

- a) Para una distribución simétrica (“normal”), el sesgo debe ser idealmente igual a “cero”
- b) Si un histograma de datos tiene una cola hacia la derecha se le llama una distribución sesgada positivamente ( $S_k > 0$ ), al contrario, es una distribución sesgada negativamente ( $S_k < 0$ ) [62].

Ejemplo 4.3. En la **Tabla 4.4** se muestra los resultados del tercer momento para los meses de enero y febrero, como podemos ver en ambos casos el resultado es mayor a cero, por lo que podemos deducir, que estas muestras presentarán una cola cargada ligeramente hacia la derecha.

**Tabla 4. 4.** Coeficiente de asimetría para el mes de enero y febrero datos muestrales de radiación

Coeficiente de asimetría (Sesgo)	
Enero	Febrero
0.178	0.176

#### 4.3.2. Cuarto momento (curtosis).

Cuarto momento (curtosis), es una medida de exceso o la escasez de datos en centro, llamado curtosis, al igual que el tercer momento es adimensional y se define como:

$$k = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} \quad (4.5)$$

Para una muestra normal, este coeficiente es idealmente cercano a 3. La presencia de un exceso de los datos en el centro de la distribución hace que  $k > 3$ . Un déficit en los datos del centro del histograma hace,  $k < 3$  [59] [62].

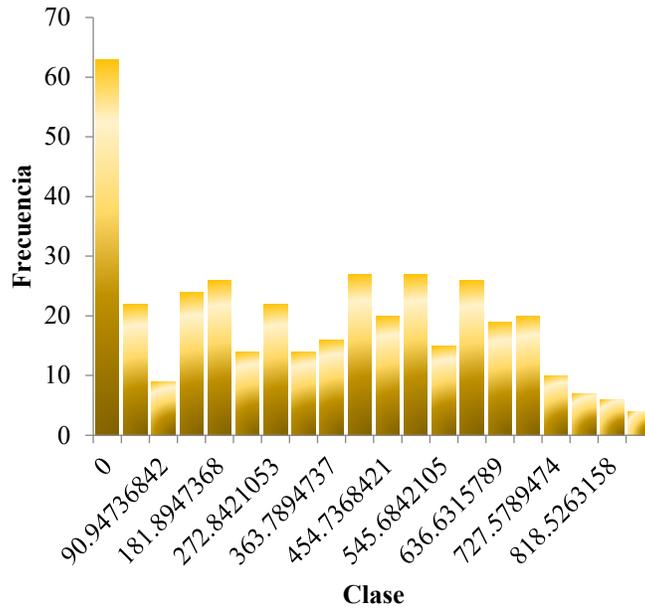
Ejemplo 4.4 la **Tabla 4. 5** muestra el cálculo de la curtosis para datos muestrales de radiación solar obtenidos de los meses de enero y febrero, como se puede ver, los valores de la curtosis en ambos casos, es menor que cero por lo que esperaríamos que el histograma para estas dos muestras presentara un déficit de datos en el centro (un histograma más aplastado hacia abajo en dirección al eje de las abscisas).

**Tabla 4. 5.** Calculo de la curtosis para las muestras de radiación meses de enero y febrero

Curtosis (cuarto momento)	
Enero	Febrero
-1.179	-1.289

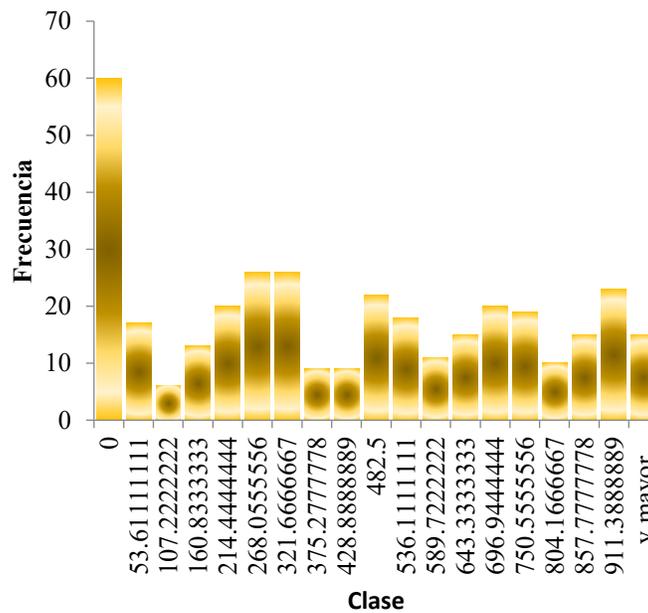
Ejemplo 3.5. Histogramas. Las **Figura 4. 3** y **Figura 4. 2** muestran los histogramas para los meses de enero y de febrero.

**Histograma mes de enero**



**Figura 4. 3.** Histograma del mes de enero

**Histograma mes de febrero**



**Figura 4. 2.** Histograma mes de febrero

# Capítulo

## 5

### Técnicas econométricas

#### Conceptos básicos.

*En primer lugar*, el trabajo empírico basado en series de tiempo supone que la serie de tiempo en cuestión es **estacionaria**.

*En segundo lugar*, se analizan varias de las causas de la autocorrelación. A veces la autocorrelación se origina porque las series de tiempo subyacentes no son estacionarias.

*En tercer lugar*, al efectuar la regresión de una variable de serie de tiempo sobre otra variable de serie de tiempo con frecuencia se obtiene una  $R^2$  muy elevada (superior a 0.9) aunque no haya una relación significativa entre las dos. En ocasiones no se espera ninguna relación entre las dos variables; sin embargo, una regresión de una variable sobre la otra a menudo muestra una relación significativa. Esta situación ejemplifica el problema de la **regresión espuria**, o **disparatada**, cuya naturaleza analizaremos en breve. Por consiguiente, es muy importante averiguar si la relación entre las variables económicas es verdadera o espuria.

*En cuarto lugar*, algunas series de tiempo financieras, como los precios de las acciones, muestran lo que se conoce como fenómeno de caminata aleatoria. Lo anterior significa que la mejor predicción para el precio de una acción, por ejemplo, de IBM, es igual a su precio actual más un choque puramente aleatorio (o término de error). De ser así, el pronóstico del precio de las acciones sería un ejercicio inútil.

*En quinto lugar*, los modelos de regresión que consideran series de tiempo son muy comunes para los pronósticos. En vista de lo expuesto, deseamos saber si tal pronóstico es válido cuando las series de tiempo sobre las cuales se basa no son estacionarias.

*Por último*, las pruebas de causalidad de Granger y Sims, suponen que las series de tiempo del análisis son estacionarias. Por consiguiente, deben efectuarse antes las pruebas para la estacionariedad que las de causalidad.

#### 5.1. Procesos estocásticos

Un proceso estocástico o aleatorio es una colección de variables aleatorias ordenadas en el tiempo. Si  $Y$  denota una variable aleatoria y es continua, se denota como  $Y(t)$ , pero si es discreta se expresa como  $Y_t$  [63].

##### 5.1.1. Procesos estocásticos estacionarios

Se dice que un proceso estocástico es estacionario si su media y su varianza son constantes en el tiempo y si su valor de la covarianza entre dos periodos depende sólo de la distancia o rezago entre estos dos periodos, y no del tiempo en el cual se calculó la covarianza. Un proceso estocástico como éste se conoce como proceso estocástico débilmente estacionario,

estacionario covariante, estacionario de segundo orden o proceso estocástico en amplio sentido [63].

Para explicar la estacionalidad débil, sea  $Y_t$  una serie de tiempo estocástica con las propiedades que se listan en las (Ecuaciones 5.1 – 5.3):

$$E(Y_t) = \mu \quad (5.1)$$

$$\text{var}(Y_t) = E(Y_t - \mu)^2 = \sigma^2 \quad (5.2)$$

$$\gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)] \quad (5.3)$$

Donde  $\gamma_k$ , la covarianza (o auto-covarianza) en el rezago  $k$ , es la covarianza entre los valores de  $Y_t$  y  $Y_{t+k}$ , es decir, entre dos valores  $Y$  separados  $k$  periodos. Si  $k = 0$ , obtenemos  $\gamma_0$ , que es simplemente la varianza de  $Y (= \sigma^2)$ ; si  $k = 1$ ,  $\gamma_1$  es la covarianza entre los dos valores adyacentes de  $Y$  [63].

### Proceso puramente aleatorio o de ruido blanco.

Se dice que un proceso es puramente aleatorio si tiene una media igual a cero, una varianza constante  $\sigma^2$  y no esta serialmente correlacionada. Supusimos que el término de error  $u_t$  que entra en el modelo clásico de regresión lineal – era un proceso de ruido blanco denotado por  $u_t \sim IIDN(0, \sigma^2)$ ; es decir,  $u_t$  está independientemente e idénticamente distribuido como una distribución normal con media cero y varianza constante. Este proceso se conoce como proceso gaussiano de ruido blanco [63].

### Procesos estocásticos no estacionarios

Aunque nuestro interés se centra en las series de tiempo estacionarias, a menudo se topa uno con series de tiempo no estacionarias, cuyo ejemplo clásico es el modelo de caminata aleatoria (MCA). Hay dos tipos de caminatas aleatorias: 1) caminata aleatoria sin deriva o sin desvío (es decir, sin término constante o de intercepto), y 2) caminata aleatoria con deriva o con desvío (es decir, hay un término constante) [63].

#### *Caminata aleatoria sin deriva*

Suponga que  $u_t$  es un término de error de ruido blanco, con media 0 y varianza  $\sigma^2$ . Entonces decimos que la serie  $Y_t$  es una caminata aleatoria si

$$Y_t = Y_{t-1} + u_t \quad (5.4)$$

En el modelo de caminata aleatoria, como se ve en la Ecuación 5.4, el valor de  $Y$  en el tiempo  $t$  es igual a su valor en el tiempo  $(t - 1)$  más un choque aleatorio; por tanto, es un modelo AR(1). Ahora bien, de 5.4, podemos escribir

$$\begin{aligned}
 Y_1 &= Y_0 + u_1 \\
 Y_2 &= Y_1 + u_2 = Y_0 + u_1 + u_2 \\
 Y_3 &= Y_2 + u_3 = Y_0 + u_1 + u_2 + u_3
 \end{aligned}$$

En general, si el proceso comenzó en el tiempo 0 con un valor de  $Y_0$ , tenemos

$$Y_t = Y_0 + \sum u_t \quad (5.5)$$

Por tanto,

$$E(Y_t) = E(Y_0 + \sum u_t) = Y_0 \quad (5.6)$$

De igual forma se demuestra que

$$var(Y_t) = t\sigma^2 \quad (5.7)$$

La suma  $\sum u_t$  se conoce también como **tendencia estocástica**.

#### *Caminata aleatoria con deriva*

Sí se modifica la [ecuación 5.4](#) de la siguiente forma:

$$Y_t = \delta + Y_{t-1} + u_t \quad (5.8)$$

Donde  $\delta$  se conoce como el **parámetro de deriva**. El término **deriva** proviene del hecho de que si escribimos la ecuación anterior como

$$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t \quad (5.9)$$

Se demuestra que  $Y_t$  se deriva o desvía hacia arriba o hacia abajo, según  $\delta$  sea positiva o negativa [63].

Observe que el modelo (5.8) también es un modelo AR(1).

Según el procedimiento analizado en la caminata aleatoria sin deriva, podemos demostrar que, para el modelo de caminata aleatoria con deriva (5.8),

$$E(Y_t) = Y_0 + t \cdot \delta \quad (5.10)$$

$$var(Y_t) = t\sigma^2 \quad (5.11)$$

Como se puede observar, para el MCA con deriva, la media, al igual que la varianza, se incrementa con el tiempo, lo que viola de nuevo las condiciones de la estacionariedad (débil). En resumen, el MCA, con o sin deriva, es un proceso estocástico no estacionario [63].

### 5.1.2. Proceso estocástico de raíz unitaria

Escribimos el MCA (5.4) como:

$$Y_t = \rho Y_{t-1} + u_t \quad -1 \leq \rho \leq 1 \quad (5.12)$$

Este modelo se parece al modelo autorregresivo de primer orden de Markov. Si  $\rho = 1$ , (5.12) se convierte en un MCA (sin deriva). Si  $\rho$  es en efecto 1, tenemos lo que se conoce como problema de raíz unitaria; es decir, enfrentamos una situación de no estacionariedad. Ya sabemos que en este caso la varianza de  $Y_t$  es no estacionaria. El nombre de raíz unitaria se debe a que  $\rho = 1$ . Por tanto, los términos no estacionariedad, caminata aleatoria, raíz unitaria y tendencia estocástica se consideran sinónimos [63].

Sin embargo, si  $|\rho| < 1$ , es decir, si el valor absoluto de  $\rho$  es menor que 1, podemos demostrar que la serie de tiempo es estacionaria. Procesos estocásticos estacionarios en tendencia (ET) y estacionarios en diferencias (ED) [63].

La distinción entre procesos estocásticos (o serie de tiempo) estacionarios y no estacionarios tiene una importancia fundamental para saber si la tendencia (la lenta evolución de largo plazo de la serie de tiempo en consideración). En términos generales, si la tendencia de una serie de tiempo es del todo predecible y no variable, se le llama **tendencia determinista**; si no es predecible, se le llama **tendencia estocástica**. Para formalizar la definición, considere el siguiente modelo de la serie de tiempo  $Y_t$ :

$$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + u_t \quad (5.13)$$

Donde  $u_t$  es un término de error de ruido blanco y donde  $t$  es el tiempo medio cronológicamente. Ahora tenemos las siguientes probabilidades:

**Caminata aleatoria pura:** Si en (5.13)  $\beta_1 = 0, \beta_2 = 0, \beta_3 = 1$ , obtenemos

$$Y_t = Y_{t-1} + u_t \quad (5.14)$$

Que no es otra cosa sino el MCA sin deriva y por lo tanto es no estacionario. Pero observe que si expresamos (5.14) como

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \quad (5.8)$$

Se convierte en estacionaria, como ya mencionamos. Por tanto, un MCA sin deriva es un **proceso estacionario en diferencias (PED)**. **Caminata aleatoria con deriva:** Si en (5.13)  $\beta_1 \neq 0$ ,  $\beta_2 = 0$ ,  $\beta_3 = 1$ , obtenemos [63].

$$Y_t = \beta_1 + Y_{t-1} + u_t \quad (5.15)$$

Que es una caminata aleatoria con deriva y en consecuencia es no estacionaria. Si la expresamos como

$$(Y_t - Y_{t-1}) = \Delta Y_t = \beta_1 + u_t \quad (5.15a)$$

Esto significa que  $Y_t$  mostrará una tendencia positiva ( $\beta_1 > 0$ ) o negativa ( $\beta_1 < 0$ ). Tal tendencia se llama **tendencia estocástica**. La ecuación (5.15a) es un PED porque la estacionariedad en  $Y_t$  se elimina al tomar las primeras diferencias de las series de tiempo.

**Tendencia determinista:** Si en (5.13),  $\beta_1 \neq 0$ ,  $\beta_2 \neq 0$ ,  $\beta_3 = 0$ , obtenemos

$$Y_t = \beta_1 + \beta_2 t + u_t \quad (5.16)$$

Lo cual se llama **proceso estacionario en tendencia (PET)**. Aunque la media de  $Y_t$  es  $\beta_1 + \beta_2 t$  no es constante, su varianza ( $= \sigma^2$ ) sí lo es. Una vez que conocemos los valores de  $\beta_1$  y  $\beta_2$ , podemos pronosticar la media sin ningún problema. Por tanto, si restamos la media de  $Y_t$  de  $Y_t$  la serie resultante será estacionaria; de ahí el nombre de **estacionario en tendencia**.

Este procedimiento de eliminar la tendencia (determinista) se llama **supresión de tendencia**.

**Caminata aleatoria con deriva y tendencia determinista:** Si en (5.13)  $\beta_1 \neq 0$ ,  $\beta_2 \neq 0$ ,  $\beta_3 = 1$ , obtenemos

$$Y_t = \beta_1 + \beta_2 t + Y_{t-1} + u_t \quad (5.17)$$

En cuyo caso tenemos una caminata aleatoria con deriva y tendencia determinista, lo cual se aprecia si expresamos esta ecuación como

$$\Delta Y_t = \beta_1 + \beta_2 t + u_t \quad (5.17a)$$

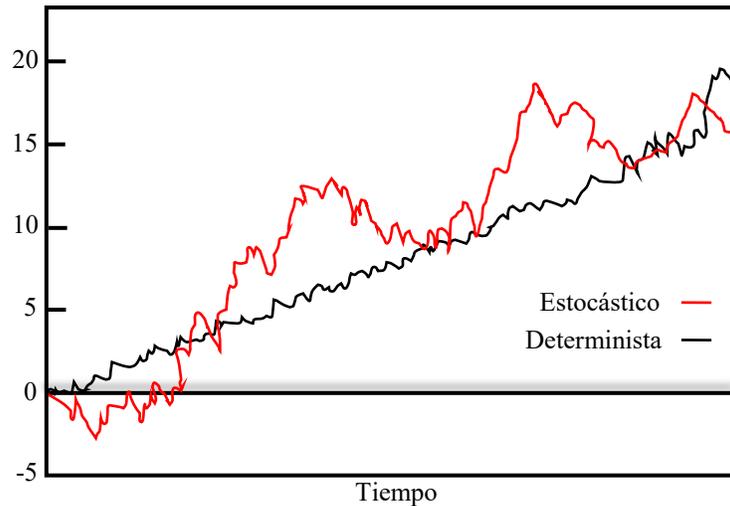
Que significa que  $Y_t$  es no estacionaria. **Tendencia determinista con componente estacionario AR(1):** Si en (5.13)  $\beta_1 \neq 0$ ,  $\beta_2 \neq 0$ ,  $\beta_3 < 1$ , tenemos

$$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + u_t \quad (5.18)$$

Que es estacionaria alrededor de la tendencia determinística. Para apreciar la diferencia entre una tendencia determinista y una estocástica, considere la (Figura 4.1) la serie llamada estocástica en esta figura está generada por el MCA con deriva:  $Y_t = 0.5 + Y_{t-1} + u_t$ , donde

se generaron 500 valores de  $u_t$  a partir de la distribución estándar y donde el valor inicial de  $Y$  se estableció como 1. La serie llamada determinista se genera de la siguiente forma:  $Y_t = 0.5t + u_t$ , donde  $u_t$  se generó como antes y  $t$  es el tiempo medio cronológicamente [63].

Como se ve a partir de la **Figura 5.1**, en el caso de la tendencia determinista, las desviaciones de la línea de tendencia (que representa la media no estacionaria) son puramente aleatorias y se eliminan rápido; no contribuyen al desarrollo de largo plazo de las series de tiempo, el cual está determinado por el componente de la tendencia  $0.5t$ . En el caso de la tendencia estocástica, por otra parte, el componente aleatorio  $u_t$  afecta el curso de largo plazo de la serie  $Y_t$ [63].



**Figura 5.1.** Tendencia determinista frente a tendencia estocástica

### 5.1.3. Procesos estocásticos integrados

El modelo de caminata aleatoria no es más que un caso específico de una clase más general de procesos estocásticos conocidos como procesos integrados. Recuerde que el MCA sin deriva es no estacionario, pero su serie de primeras diferencias, como se muestra en (5.13), es estacionaria.

Por tanto, el MCA sin deriva se llama proceso integrado de orden 1 y se denota como  $I(1)$ . De manera similar, si una serie de tiempo tiene que diferenciarse dos veces (es decir, se toman primeras diferencias de la serie de primeras diferencias) para hacerla estacionaria, esa serie de tiempo se denomina integrada de orden 2. En general, si una serie de tiempo (no estacionaria) debe diferenciarse  $d$  veces para hacerla estacionaria, decimos que la serie es integrada de orden  $d$ . Una serie de tiempo  $Y_t$  integrada de orden  $d$  se denota como  $Y_t \sim I(d)$ . Si una serie de tiempo es estacionaria desde el principio (es decir, si no requiere ninguna diferenciación), decimos que es integrada de orden cero y se denota mediante  $Y_t \sim I(0)$ . Por tanto, con los términos “serie de tiempo estacionaria” y “serie de tiempo integrada de orden cero” daremos a entender la misma cosa [63].

La mayoría de las series de tiempo económicas son  $I(1)$ ; es decir, por lo general se convierten en estacionarias sólo después de tomar sus primeras diferencias [63].

### El fenómeno de regresión espuria

Para ver por qué las series de tiempo estacionarias son tan importantes, considere los dos modelos de caminata aleatoria siguientes:

$$Y_t = Y_{t-1} + u_t \quad (5.19)$$

$$X_t = X_{t-1} + v_t \quad (5.20)$$

donde se generaron 500 observaciones de  $u_t$  a partir de  $u_t \sim N(0, 1)$  y 500 observaciones de  $v_t$  a partir de  $v_t \sim N(0, 1)$ , además de que se supuso que los valores iniciales de  $Y$  y  $X$  eran cero. También se supuso que  $u_t$  y  $v_t$  no están serial ni mutuamente correlacionadas. Como ya sabemos, ambas series de tiempo son no estacionarias; es decir, son  $I(1)$  o exhiben tendencias estocásticas [63].

### 5.2. Significancia estadística de los coeficientes de autocorrelación

La significancia estadística de cualquier  $\hat{\rho}_k$  se juzga mediante el error estándar. Bartlett demostró que, si una serie de tiempo es puramente aleatoria, es decir, si es una muestra de ruido blanco, los coeficientes de autocorrelación muestrales  $\hat{\rho}_k$  son aproximadamente [63].

$$\hat{\rho}_k \sim N(0, 1/n) \quad (5.21)$$

Es decir, en muestras grandes, los coeficientes de autocorrelación muestrales están normalmente distribuidos y tienen media cero y varianza igual a  $1/n$  sobre el tamaño de la muestra. Es decir, si se tuvieran 244 observaciones, la varianza sería  $1/244 \approx 0.0041$ , y el error estándar,  $\sqrt{0.0041} \approx 0.0640$ . Por tanto, según las propiedades de la distribución normal estándar, el intervalo de confianza de 95% para cualquier (población)  $\rho_k$  es:

$$\hat{\rho} \pm 1.96(0.0640) = \hat{\rho}_k \pm 0.1254 \quad (5.22)$$

En otras palabras,

$$Prob(\hat{\rho}_k - 0.1254 \leq \rho_k \leq \hat{\rho}_k + 0.1254) = 0.95 \quad (5.23)$$

Si el intervalo anterior incluye el valor cero, no rechazamos la hipótesis de que la verdadera  $\rho_k$  es cero, pero si este intervalo no incluye 0, rechazamos la hipótesis de que la verdadera  $\rho_k$  es cero [63] [64].

En lugar de probar la significancia estadística de cualquier coeficiente de autocorrelación individual, para probar la hipótesis conjunta de que todos los  $\rho_k$  hasta ciertos rezagos son simultáneamente iguales a cero, podemos utilizar el **estadístico Q** desarrollado por Box y Pierce, que se define como

$$Q = n \sum_{k=1}^m \hat{\rho}_k^2 \quad (5.24)$$

Donde  $n$  = tamaño de la muestra y  $m$  = longitud del rezago. El estadístico  $Q$  es común para probar si una serie de tiempo es de ruido blanco. En muestras grandes, este estadístico se distribuye aproximadamente como la distribución  $\chi^2$  con  $m$  gl. En una aplicación, si la  $Q$  calculada excede el valor  $Q$  crítico de la distribución  $\chi^2$  en el nivel de significancia seleccionado, podemos rechazar la hipótesis nula de que todos los  $\rho_k$  (verdaderos) son iguales a cero; por lo menos algunos de ellos deben ser diferentes de cero [63][65].

Una variante del estadístico  $Q$  de Box-Pierce es el **estadístico Ljung-Box (LB)**, que se define como

$$LB = n(n+2) \sum_{k=1}^m \left( \frac{\hat{\rho}_k^2}{n-k} \right) \sim \chi^2 m \quad (5.25)$$

Aunque en muestras grandes tanto el estadístico  $Q$  como el estadístico  $LB$  siguen la distribución  $\chi^2$  con  $m$  gl, se ha visto que el estadístico  $LB$  tiene mejores propiedades en muestras pequeñas que el estadístico  $Q$  [66].

### 5.3. Prueba de raíz unitaria

Otra prueba sobre estacionariedad (o no estacionariedad) que se populariza cada vez más se conoce como **prueba de raíz unitaria**. Primero la explicaremos, luego la ilustraremos y después consideraremos algunas limitantes de esta prueba [65].

El punto de partida es el proceso (estocástico) de raíz unitaria, se inicia con

$$Y_t = \rho Y_{t-1} + u_t \quad -1 \leq \rho \leq 1 \quad (5.12)$$

Donde  $u_t$  es un término de error de ruido blanco.

Sabemos que si  $\rho = 1$ , es decir, en el caso de la raíz unitaria, 5.12 se convierte en un modelo de caminata aleatoria sin deriva, del cual sabemos también que es un proceso estocástico no estacionario. Por consiguiente, ¿por qué no simplemente hacer la regresión de  $Y_t$  sobre su valor rezagado (de un periodo)  $Y_{t-1}$  y se averigua si la  $\rho$  es estadísticamente igual a 1? De ser así,  $Y_t$  es no estacionaria. Esta es la idea general de la prueba de raíz unitaria para la estacionariedad [63].

Sin embargo, no podemos estimar la ecuación [5.12](#) por MCO y probar la hipótesis de que  $\rho = 1$  por medio de la prueba  $t$  acostumbrada, porque esa prueba tiene un sesgo muy marcado. Por tanto, manipulamos [5.12](#) de la siguiente forma: restamos  $Y_{t-1}$  de ambos miembros de la ecuación [5.12](#) para obtener:

$$\begin{aligned} Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + u_t \\ &= (\rho - 1)Y_{t-1} + u_t \end{aligned} \quad (5.26)$$

La cual también se expresa como:

$$\Delta Y_t = \delta Y_{t-1} + u_t \quad (5.27)$$

Donde  $\delta = (\rho - 1)$  y  $\Delta$ , como siempre, es el operador de primeras diferencias. Por tanto, en la práctica, en vez de estimar [\(5.12\)](#), calculamos [\(5.27\)](#) y probamos la hipótesis (nula) de que  $\delta = 0$ , y la hipótesis alternativa es que  $\delta < 0$ . Si  $\delta = 0$ , entonces  $\rho = 1$ ; es decir, tenemos una raíz unitaria, lo cual significa que la serie de tiempo en consideración es no estacionaria. Antes de proceder con la estimación [\(5.27\)](#) debemos observar que si  $\delta = 0$ , entonces [5.27](#) se convertirá en

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \quad (5.29)$$

Como  $u_t$  es un término de error de ruido blanco, entonces es estacionario, lo cual significa que las primeras diferencias de una serie de tiempo de caminata aleatoria son estacionarias. Se realiza la regresión sobre  $Y_{t-1}$ , a fin de ver si el coeficiente estimado de la pendiente en esta regresión ( $= \hat{\delta}$ ) es o no cero. Si es cero, concluimos que  $Y_t$  es no estacionaria; pero si es negativa, se infiere que  $Y_t$  es estacionaria. La única interrogante es saber con qué prueba averiguar si el coeficiente estimado de  $Y_{t-1}$  en [\(5.27\)](#) es o no cero. Uno estaría tentado a utilizar la prueba  $t$  usual. Por desgracia, según la hipótesis nula de que  $\delta = 0$  (es decir,  $\rho = 1$ ), el valor  $t$  del coeficiente estimado de  $Y_{t-1}$  no sigue la distribución  $t$  ni siquiera en muestras grandes, es decir, no tiene una distribución normal asintótica [\[63\]](#).

Dickey y Fuller probaron que según la hipótesis nula de que  $\delta = 0$ , el valor estimado  $t$  del coeficiente  $Y_{t-1}$  en [\(5.27\)](#) sigue el estadístico  $\tau$ . Estos autores calcularon los valores críticos del estadístico  $\tau$  con base en simulaciones Monte Carlo. En la bibliografía, el estadístico o prueba  $\tau$  se conoce como prueba Dickey – Fuller (DF) [\[63\]](#).

Al fin de permitir las distintas posibilidades, la prueba DF se estima en tres diferentes formas, es decir, conforme a tres hipótesis nulas:

$$Y_t \text{ es una caminata aleatoria:} \quad \Delta Y_t = \delta Y_{t-1} + u_t \quad (5.27)$$

$$Y_t \text{ es una caminata aleatoria con deriva:} \quad \Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t \quad (5.29)$$

$$Y_t \text{ es una caminata aleatoria con deriva} \\ \text{alrededor de una tendencia} \quad \Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t \quad (5.30)$$

determinista:

donde  $t$  es la variable de tiempo o de tendencia. En cada caso, las hipótesis son:

*Hipótesis nula:*  $H_0: \delta = 0$  (es decir, existe una raíz unitaria, la serie de tiempo es no estacionaria o tiene tendencia estocástica) [63].

*Hipótesis alternativa:*  $H_1: \delta < 0$  (es decir, la serie de tiempo es estacionaria, posiblemente alrededor de una tendencia determinista) [63].

Si rechazamos la hipótesis nula, esto significa que 1)  $Y_t$  es estacionaria con media cero en el caso de la ecuación (5.27) o que 2)  $Y_t$  es estacionaria con una media distinta de cero en el caso de (5.29) en el caso de la ecuación (5.30), podemos probar que  $\delta < 0$  (es decir, no hay tendencia estocástica) y  $\alpha \neq 0$  (es decir, la existencia de una tendencia determinista) simultáneamente, mediante la prueba  $F$  pero con valores críticos tabulados por Dickey y Fuller. Cabe señalar que una serie de tiempo puede contener tanto una tendencia estocástica como una determinista [63].

Si el valor absoluto calculado del estadístico ( $|\tau|$ ) excede la DF absoluta o los valores críticos  $\tau$  de MacKinnon, rechazamos la hipótesis de que  $\delta = 0$ , en cuyo caso la serie de tiempo es estacionaria. Por otra parte, si el  $|\tau|$  no excede el valor crítico  $\tau$ , no rechazamos la hipótesis nula, en cuyo caso la serie de tiempo es no estacionaria. En la mayoría de las aplicaciones, el valor  $\tau$  es negativo. Por consiguiente, también vale decir que si el valor  $\tau$  calculado (negativo) es más pequeño (es decir, más negativo) que el valor crítico  $\tau$ , rechazamos la hipótesis nula (es decir, la serie de tiempo es estacionaria); de lo contrario, no la rechazamos (es decir, la serie de tiempo es no estacionaria).

### 5.3.1. La prueba Dickey-Fuller aumentada (DFA)

Al llevar a cabo la prueba DF en (5.27), (5.29) o (5.30) supusimos que el término de error  $u_t$  no estaba correlacionado. Pero Dickey y Fuller desarrollaron una prueba cuando dicho término sí está correlacionado, la cual se conoce como **prueba Dicky – Fuller aumentada (DFA)**. Esta prueba implica “aumentar” las tres ecuaciones anteriores mediante la adición de los valores rezagados de la variable dependiente  $\Delta Y_t$ . Para un ejemplo específico, suponga que utilizamos (5.30). La prueba DFA consiste en este caso en estimar la siguiente regresión:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{(i=1)}^m \alpha_i \Delta Y_{t-i} + \varepsilon_t \quad (4.31)$$

Donde  $\varepsilon_t$  es un término de error puro de ruido blanco y donde  $\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2})$ ,  $\Delta Y_{t-2} = (Y_{t-2} - Y_{t-3})$ , etc. El número de términos de diferencia rezagados que debemos incluir con frecuencia se determina de manera empírica, con la idea de incluir los términos suficientes para que el término de error en (5.31) no esté serialmente relacionado y sea posible obtener una estimación insesgada de  $\delta$ , el coeficiente  $Y_{t-1}$  rezagado [63].

#### 5.4. Modelos autorregresivos y de rezagos distribuidos.

En un análisis de series de tiempo, cuando el modelo de regresión incluye no solo los valores actuales sino valores rezagados (pasados) de las variables explicativas (las  $X$ ), se denominan modelos de rezagos distribuidos, si el modelo incluye uno o más valores de la variable dependiente entre sus variables explicativas, se denomina modelo autorregresivo. Así,

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t \quad (5.32)$$

representa un modelo de rezagos distribuidos, mientras que

$$Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + u_t \quad (5.33)$$

es un ejemplo de modelo autorregresivo. Dónde  $\partial Y_t / \partial X_{t-k} = \beta_k$ .

En economía, la dependencia de una variable  $Y$  (la variable dependiente) respecto de otra u otras variables  $X$  (las variables explicativas) pocas veces es instantánea. Con frecuencia  $Y$  responde a  $X$  en un lapso, el cual se denomina rezago [63].

En este tipo de modelos podemos tener un rezago finito de  $k$  periodos y el coeficiente  $\beta_0$  se conoce como multiplicador de corto plazo o de impacto porque da el cambio en el valor medio de  $Y$  que sigue a un cambio unitario en  $X$  en el mismo periodo. Si el cambio en  $X$  se mantiene igual desde el principio, entonces  $(\beta_0 + \beta_1)$  da el cambio en (el valor medio de)  $Y$  en el periodo siguiente  $(\beta_0 + \beta_1 + \beta_2)$  en el que le sigue, y así sucesivamente. Estas sumas parciales se denominan multiplicadores ínterin, o intermedios. Por último, después de  $k$  periodos obtenemos

$$\sum_{i=0}^k \beta_i = \beta_0 + \beta_1 + \beta_2 + \dots + \beta_k = \beta \quad (5.34)$$

que se conoce como multiplicador de rezagos distribuidos de largo plazo o total, siempre que exista la suma  $\beta$  [63].

Si definimos

$$\beta_i^* = \frac{\beta_i}{\sum \beta_i} = \frac{\beta_i}{\beta} \quad (5.35)$$

obtenemos  $\beta_i^*$  “estandarizado”. Las sumas parciales del  $\beta_i^*$  estandarizado dan la proporción del impacto de largo plazo, o total, sentido durante cierto periodo [63].

## 5.5. Prueba de cointegración de Eagle-Granger

El tiempo no corre hacia atrás. Es decir, si un acontecimiento  $A$  sucede antes de un suceso  $B$ , es posible que  $A$  cause  $B$ . Sin embargo, no es posible que  $B$  provoque  $A$ . En otras palabras, los acontecimientos pasados pueden propiciar sucesos que ocurren en la actualidad. Lo cual no sucede con los sucesos futuros [63].

El enunciado “ $y_i$  causa  $y_j$ ” es sólo una forma abreviada del enunciado más preciso, aunque más extenso de que “ $y_i$  contiene información útil para predecir  $y_j$ ” [63].

### 5.5.1. Prueba de Granger

Para explicar la prueba de Granger consideraremos una pregunta frecuente en macroeconomía: ¿el  $PIB$  “causa” la oferta de dinero  $M$  ( $PIB \rightarrow M$ ), o la oferta de dinero  $M$  causa el  $PIB$  ( $M \rightarrow PIB$ )? (La flecha señala la dirección de la causalidad.) La prueba de causalidad de Granger supone que la información relevante para la predicción de las variables respectivas,  $PIB$  y  $M$ , está contenida únicamente en la información de series de tiempo sobre estas variables. La prueba implica la estimación de las siguientes regresiones:

$$PIB_t = \sum_{i=1}^n \alpha_i M_{t-i} + \sum_{j=1}^n \beta_j PIB_{t-j} + u_{1t} \quad (5.36)$$

$$M_t = \sum_{i=1}^n \lambda_i M_{t-i} + \sum_{j=1}^n \delta_j PIB_{t-j} + u_{2t} \quad (5.37)$$

donde se supone que las perturbaciones  $u_{1t}$  y  $u_{2t}$  no están correlacionadas. A propósito, observe que en vista de que hay dos variables, tratamos con una causalidad bilateral [63].

De modo más general, como el futuro no puede predecir el pasado, si la variable  $X$  (a la manera de Granger) causa la variable  $Y$ , los cambios en  $X$  deben preceder a los cambios en  $Y$ . Por consiguiente, en una regresión de  $Y$  sobre otras variables (con sus propios valores pasados), si incluimos valores pasados o rezagados de  $X$  y esto mejora significativamente la predicción de  $Y$ , podemos decir que  $X$  (a la manera de Granger) causa  $Y$ . Aplica una definición similar si  $Y$  (a la manera de Granger) causa  $X$  [63].

Los pasos comprendidos en la prueba de causalidad de Granger son los siguientes [63].

1. Haga la regresión del  $PIB$  actual sobre todos los términos rezagados del  $PIB$  y otras variables, en su caso, pero no incluya las variables  $M$  rezagadas en esta regresión, ésta es la regresión restringida. A partir de esta regresión, obtenga la suma de cuadrados residuales restringida,  $SCR_R$ .
2. Ahora efectúe la regresión con los términos rezagados  $M$ . A partir de esta regresión, obtenga la suma de cuadrados residuales no restringida,  $SCR_{NR}$ .

3. La hipótesis nula es  $H_0: \alpha_i = 0, i = 1, 2, \dots, n$ , es decir, los términos rezagados de  $M$  no pertenecen a la regresión.
4. Para probar esta hipótesis, aplicamos la prueba  $F$  dada por (5.38), a saber,

$$F = \frac{(SCR_R - SCR_{NR})/m}{SCR_{NR}/(n - k)} \quad (5.38)$$

que sigue la distribución  $F$  con  $m$  y  $(n - k)$  gl. En el presente caso,  $m$  es igual al número de términos rezagados de  $M$ , y  $k$  es el número de parámetros estimados en la regresión no restringida.

5. Si el valor  $F$  calculado excede al valor  $F$  crítico en el nivel seleccionado de significancia, rechazamos la hipótesis nula, en cuyo caso los términos rezagados de  $M$  pertenecen a la regresión. Ésta es otra forma de decir que  $M$  causa al  $PIB$ .
6. Se repiten los pasos 1 a 5 para probar el modelo (5.37), es decir, para definir si el  $PIB$  causa  $M$ .

Antes de ilustrar la prueba de causalidad de Granger, hay varios puntos que deben tomarse en cuenta:

1. Se supone que las dos variables,  $PIB$  y  $M$ , son estacionarias. A veces, al tomar las primeras diferencias de las variables, las convierte en estacionarias, si es que aún no son estacionarias en la forma de nivel.
2. El número de términos rezagados por introducir en las pruebas de causalidad es una cuestión práctica importante. Como en el caso de los modelos de rezagos distribuidos, quizá tengamos que elegir mediante el criterio de información *Akaike* o *Schwarz*.
3. Supusimos que los términos de error en la prueba de causalidad no están correlacionados. De no ser así, quizá deba llevarse a cabo la transformación adecuada.
4. Como nos interesa probar la causalidad, no se requiere presentar los coeficientes estimados de los modelos (5.36) y (5.37) en forma explícita (para ahorrar espacio); bastan los resultados de la prueba  $F$  dados en (5.38).
5. Es necesario tener cuidado con la causalidad “espuria”. En el ejemplo del  $PIB$  y la oferta de dinero, suponga que consideramos la tasa de interés (por ejemplo, la de corto plazo). Es muy posible que el dinero “cause según Granger” la tasa de interés y que la tasa de interés, a su vez, “cause según Granger” el  $PIB$ . Por tanto, si no tomamos en cuenta la tasa de interés y concluimos que el dinero causa el  $PIB$ , la causalidad

observada entre el *PIB* y la oferta de dinero puede ser espuria. Como ya señalamos, una forma de resolver esto es considerar un sistema de ecuaciones múltiples, como la autorregresión vectorial (VAR) [63].

# Capítulo

## 6

### Técnicas de predicción y análisis de series de tiempo

#### 6.1. Modelos estadísticos para la predicción de series de tiempo.

##### 6.1.1. Herramientas básicas de pronóstico.

##### Series de tiempo y datos de sección transversal.

A menudo nuestros datos históricos son de una secuencia de observaciones sobre el tiempo. Llamamos a tal secuencia una serie de tiempo. Por ejemplo, cifras mensuales de ventas, precios de las acciones diarias, tasa de interés semanal, ganancias anuales, temperaturas máximas diarias, producciones de cultivo anual, y mediciones de electrocardiograma, todas son series de tiempo [67][68].

En el pronóstico, tratamos de estimar como las secuencias de observaciones continuaran en el futuro [68].

##### 6.1.2. Resúmenes gráficos

Lo más importante por hacer para la primera exploración de los datos es visualizar los datos a través de las gráficas. Algunas veces las gráficas sugieren posibles explicaciones para algunas de las variaciones en los datos. El tipo de datos determinará qué tipo de gráfica es más apropiado [68].

##### Gráficas de tiempo y patrones de series de tiempo

Para las series de tiempo, la forma de gráfica más obvia es una gráfica de tiempo. Una gráfica de tiempo inmediatamente revela cualquier tendencia sobre el tiempo, cualquier comportamiento regular de estacionalidad, y otras características sistemáticas de los datos. Estos necesitan ser identificados de tal manera que puedan ser incorporados dentro del modelo estadístico [68].

Una importante etapa en la selección de un método de pronóstico apropiado es considerar los tipos de patrones de datos, de tal manera que los métodos más apropiados para dichos patrones se puedan utilizar, se pueden distinguir cuatro tipos de patrones de datos en las series de tiempo: horizontal, estacional, cíclico, y de tendencia [68].

1. Un patrón *horizontal* (H) existe cuando los valores de los datos fluctúan alrededor de un promedio constante. (Dicha serie se llama “estacionaria” en su media).

2. Un patrón *estacional* (S) existe cuando una serie es influenciada por factores estacionales (e.g., el trimestre de un año, el mes o el día de la semana). Las series estacionales algunas veces también son llamadas “periódicas” sin embargo no se repiten exactamente así mismas en cada periodo.
3. Un patrón *cíclico* (C) existe cuando los datos exhiben alzas y caídas que no son de un periodo fijo. Para series económicas, estas usualmente se deben a fluctuaciones económicas tales como aquellas asociadas al ciclo comercial. La mayor distinción entre un patrón estacional y uno cíclico es que el primero es de una longitud constante y se repite sobre una base periódica regular, mientras que el último varía en longitud, por otra parte, la longitud promedio de un ciclo es usualmente más larga que aquella de estacionalidad y la magnitud de un ciclo es usualmente más variable que aquella de estacionalidad.
4. Un patrón de *tendencia* (T) existe cuando hay un incremento o decremento a largo plazo en los datos. Una de las cosas que hacen un pronóstico interesante y desafiante es la amplia variedad de patrones que comúnmente suceden en datos de series reales de tiempo.

### 6.1.3. Resúmenes numéricos.

Además de las gráficas, es también útil proporcionar resúmenes numéricos, un resumen numérico para un conjunto de datos se llama un *estadístico*. Para un conjunto de datos únicos (datos univariados) o una serie de tiempo simple, los estadísticos descriptivos más comunes son la media, la desviación estándar y la varianza [68].

Para un par de variables aleatorias (datos bivariados) es de interés describir como los dos conjuntos de datos se relacionan uno con el otro. El resumen numérico más ampliamente utilizado (estadístico) para estos propósitos son la covarianza y la correlación [68].

A continuación, para una sola serie de tiempo, es útil comparar la observación en un periodo de tiempo con la observación en otro periodo de tiempo. Los dos estadísticos aquí más comunes son la auto-covarianza y la auto-correlación [68].

### Estadística univariante.

Considere la radiación solar que consta de 19 mediciones dados en la (Tabla 5.1), Los valores de radiación solar han sido numerados del 1 al 19 para una fácil referencia.

Utilizando las letras *RS* para denotar la radiación y un subíndice  $i$  ( $i = 1, 2, 3, \dots, 19$ ) para denotar el  $i$ -ésimo dato de radiación solar, la radiación solar media puede ser escrito

$$\begin{aligned}\bar{M} &= (M_1 + M_2 + M_3 + \dots + M_{19})/19 \\ &= \frac{1}{19} \sum_{i=1}^{19} M_i\end{aligned}$$

$$\frac{10151}{19} = 534.3 \text{ W/m}^2$$

La media no se debe confundir con la *mediana*, que es la observación de la mitad. Así para los 19 datos anteriores, la mediana en  $\text{W/m}^2$  es el décimo dato cuando se acomodan en orden ascendente. De esta forma, la mediana de los datos es de  $591 \text{ W/m}^2$ . Tanto la media como la mediana están diseñadas para proporcionar una medida numérica del centro del conjunto de los datos [68].

**Tabla 6. 1.** Radiación solar correspondiente a enero y febrero

	Enero	Febrero
1	177	174
2	393	396
3	583	593
4	724	745
5	791	820
6	784	837
7	707	770
8	563	640
9	369	457
10	156	232
11	5	3
12	185	135
13	409	334
14	605	466
15	745	598
16	816	661
17	812	662
18	736	610
19	591	489

Además de medir el centro del conjunto de los datos, también es valioso medir la dispersión de los datos. Esto es, queremos una medida numérica que nos indique si los datos están estrechamente agrupados o dispersos a través de un amplio rango, los temas de centro de datos y dispersión se trataron más fondo en el (**Capítulo 4**).

### Estadística bivariada

Cuando se grafican el precio de un automóvil versus el kilometraje, podría parecer que existe una relación negativa entre estas dos variables. Por relación negativa nos referimos que conforme el kilometraje aumenta, el precio del automóvil tiende a sufrir un decremento. (Una relación positiva sería similar a la relación de altura versus peso – conforme incrementa la altura también lo hace el peso.)

Como sea se trata de dos pares de observaciones (e.g., precio y kilometraje, altura y peso, precio y demanda), es de interés examinar y medir la medida de la relación entre las dos variables [69].

Suponga que denotamos las dos variables por  $X$  y  $Y$ . Un estadístico que indica como dos variables “co-varían” se llama *covarianza* y se define como sigue:

$$Cov_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (6.1)$$

Donde  $\bar{X}$  y  $\bar{Y}$  son los promedios de  $X$  y  $Y$ , respectivamente, y  $n$  es el número de observaciones en cada variable [69].

Las unidades de la covarianza son problemáticas. Es difícil interpretar miles de dólares – millas por galón. Por lo tanto, el valor de los cálculos del coeficiente de correlación, se describe a continuación. Tenga en cuenta que el valor de la covarianza entre el precio y el kilometraje es negativo, pero la magnitud de la covarianza  $Cov_{PM}$  claramente depende de las unidades involucradas. Si las cifras del kilometraje fueran en kilometro por litro y el precio en dólares, la gráfica luciría similar pero la covarianza sería bastante diferente [69].

El *coeficiente de correlación*, designado como  $r$ , es una medida especial de la covarianza que se ocupa del problema de escala que se acaba de mencionar. Si la covarianza ( $Cov_{XY}$ ) es dividida por las dos desviaciones estándar ( $S_X$  y  $S_Y$ ), entonces las unidades en el numerador y en el denominador se cancelan, quedando un número adimensional, que es el coeficiente de correlación entre  $X$  y  $Y$ . Esto se escribe como sigue:

$$r_{XY} = \frac{Cov_{XY}}{S_X S_Y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}} \quad (6.2)$$

El efecto de esta escala (dividir la  $Cov_{XY}$  por  $S_X$  y  $S_Y$ ) es restringir el rango de  $r_{XY}$  a un intervalo de  $-1$  a  $+1$ . No importan las unidades de medición para  $X$  y  $Y$  el coeficiente de correlación,  $r_{XY}$ , siempre se restringe para hallarse en este intervalo [69].

Este número resultante es fácilmente interpretable. Por ejemplo, si tuvieramos una correlación de  $-0.73$  entre el precio y el kilometraje, sería una correlación negativa y sustancial. Existe una fuerte asociación negativa entre el precio y el kilometraje [69].

La covarianza y especialmente la correlación son estadísticos básicos para conjuntos de datos de dos variables y para más amplios conjuntos de datos multi-variantes. Se debe tener cuidado, sin embargo, en recordar que estas son medidas de asociación lineal entre dos variables, de modo que no es apropiado (significativamente) para aplicarlas a medidas de correlación cuando hay una relación curvilínea entre dos variables [69].

### 6.1.4. Autocorrelación

La covarianza y el coeficiente de correlación son estadísticos (sumarios estadísticos) que miden el grado de la relación lineal entre dos variables. Como tales se pueden utilizar para identificar relaciones explicativas. La auto-covarianza y la auto-correlación son medidas comparables que sirven para el mismo propósito para una sola serie de tiempo [66][69].

Por ejemplo, si comparamos  $Y_t$  (Observación en un tiempo  $t$ ) con  $Y_{t-1}$  (Observación en un tiempo  $t - 1$ ), entonces se ve como observaciones consecutivas se relacionan. La observación  $Y_{t-1}$  se describe como “rezago” por un periodo. Similarmente, es posible comprar observaciones rezagadas de dos periodos, tres periodos, y así sucesivamente.

La **Tabla 6. 2** muestra la serie de cerveza, que es una serie de tiempo simple de 56 meses desde enero de 1991 a agosto 1995. Las observaciones  $Y_1, Y_2, \dots, Y_{56}$  son observaciones en tiempos de periodos 1, 2, ..., 56 respectivamente. Si rezagamos la serie por un periodo, como se muestra en la columna 3, entonces habrá 55 pares de observaciones para comparar. Sin embargo, ya que son una y la misma serie (con un rezago en un periodo) la recopilación de medidas es llamada auto-covarianza y auto-correlación [66].

Ya que las dos series son casi las mismas, en lugar de utilizar la (Ecuación 6.1 y 6.2) para calcular la auto-covarianza y la auto-correlación, normalmente se utilizan las (Ecuaciones 6.3 y 6.4). Se denota la auto-covarianza de un retardo  $k$  por  $c_k$  y la auto-correlación de un retardo  $k$  por  $r_k$ . Entonces definido [66].

$$C_k = \frac{1}{n} \sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y}) \tag{6.3}$$

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \tag{6.4}$$

**Tabla 6. 2.** Cálculo de la auto-covarianza y la auto-correlación utilizando las ecuaciones 6.3 y 6.4, y un rezago de un periodo.

(1) $t$	(2) $Y_t$	(3) $Y_{t-1}$	(4) $(Y_t - \bar{Y})$	(5) $(Y_{t-1} - \bar{Y})$	(6) $(Y_t - \bar{Y})^2$	(7) $(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})$
1	164	-	14.70	-	215.99	-
2	148	164	-1.30	14.70	1.70	-19.16
3	152	148	2.70	-1.30	7.27	-3.51
4	144	152	-5.30	2.70	28.13	-14.30
5	155	144	5.70	-5.30	32.45	-30.21
6	125	155	-24.30	5.70	590.66	-138.44
⋮	⋮	⋮	⋮	⋮	⋮	⋮
53	151	127	1.70	-22.30	2.88	-37.84
54	130	151	-19.30	1.70	372.63	-32.75
55	119	130	-30.30	-19.30	918.31	584.97
56	153	119	3.70	-30.30	13.66	-112.01
Sum	8361				21135.84	8893.51

$$\text{Promedio } \bar{Y} = \frac{8361}{56} = 149.30$$

$$\text{Auto-covarianza rezago 1 } c_1 = \frac{8893.51}{56} = 158.8$$

$$\text{Autocorrelación rezago 1 } r_1 = \frac{8893.51}{21135.84} = 0.421$$

### Función de auto-correlación

Juntos, la auto-correlación y los rezagos 1, 2, ..., componen la función de auto-correlación o *FAC*. en lugar de escanear una lista de números, es mucho más fácil graficar las auto-correlaciones en contra de los rezagos. Como una gráfica se conoce como un correlograma y nos ayuda a visualizar la *FAC* rápidamente y fácilmente [68].

#### 6.1.5. Medición de la precisión de los pronósticos.

Ahora retornamos a otro asunto fundamental como medir la idoneidad de un método particular de predicción para un conjunto de datos dado. En la mayoría de las situaciones de pronóstico, la predicción se trata como el criterio primordial para seleccionar un método de predicción. En muchos casos, la palabra “precisión” se refiere al “buen ajuste,” que a su vez se refiere a cuan bien el modelo de predicción es capaz de reproducir los datos que ya se conocen. Para el consumidor de pronósticos, es la exactitud del pronóstico futuro lo que es más importante [69].

Para ilustrar los cálculos involucrados, nos referiremos a la producción mensual de cerveza en Australia. La (Tabla 6. 3) muestra los últimos 8 meses de observaciones (enero-agosto 1995). La segunda columna muestra predicciones para estos valores, obtenidos utilizando un método muy simple, al tomar los promedios de cada mes sobre los cuatro años pasados. Entonces, por ejemplo, la predicción para enero de 1995 es el promedio de producción de enero de 1991, enero de 1992, enero de 1993, y enero de 1994 [69].

Tabla 6. 3

Periodo	Observación	Pronóstico
$t$	$Y_t$	$F_t$
1	138	150.25
2	136	139.50
3	152	157.25
4	127	143.50
5	151	138.00
6	130	127.50
7	119	138.25
8	153	141.50
9	-	140.50
10	-	167.25

### Medidas estadísticas estándar.

Si  $Y_t$  es la observación actual para un periodo de  $t$  y  $F_t$  es el pronóstico para el mismo periodo, entonces el error se define como

$$e_t = Y_t - F_t \quad (6.5)$$

Usualmente,  $F_t$  se calcula utilizando los datos  $Y_1, \dots, Y_{t-1}$ . Esto es pronóstico de un paso ya que se pronostica un periodo adelanté de la última observación utilizada en los cálculos. Por lo tanto, se describe  $e_t$  error de pronóstico de un paso. Esto es la diferencia entre la observación  $Y_t$  el pronóstico hecho usando todas las observaciones, pero no incluyendo  $Y_t$  [69].

## 6.2 La metodología Box – Jenkins para modelos ARIMA.

### 6.2.1. Examen de correlaciones en los datos de series de tiempo

En esta sección nos concentraremos en cierto análisis que pueden ser aplicados a una serie de tiempo empírica para determinar sus propiedades estadísticas [66][69].

Las bases de la aproximación Box-Jenkins para el modelado de las series de tiempo se resume en la (Figura 6. 1) y se compone de tres fases: identificación, estimación y pruebas, y aplicación [66][69].

### La función de autocorrelación

La clave estadística en el análisis de las series de tiempo es el coeficiente de auto-correlación (o la correlación de las series de tiempo consigo misma, retrasos de 1,2 o más periodos). Esto se introdujo en la Sección 6.1.4. Recordando la formula [66][69].

$$r_k = \sum_{t=k+1}^n \frac{(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (6.6)$$

Entonces  $r_1$  indica cómo los valores sucesivos de  $Y$  se relacionan unos con otros,  $r_2$  indica cómo valores de  $Y$  separados dos periodos se relacionan uno con otro, y así sucesivamente. Juntas, las autocorrelaciones con rezagos  $1, 2, \dots$ , componen la *función de autocorrelación* o FAC [66][69].

Considere la serie de tiempo consistente de 36 observaciones en la (Tabla 6. 4) y graficada en la (Figura 6. 2). Esta serie se construye utilizando números aleatorios no correlacionados entre 0 y 100. De cualquier manera, se supone, que este hecho no se conoce, se puede determinar al aplicar análisis de autocorrelación. Para datos no correlacionados, esperaríamos

que cada autocorrelación fuera cercana a cero. La (Figura 6. 3) muestra los coeficientes de autocorrelación para los datos en la (Tabla 6. 4 ), para tiempos de rezago de 1,2,3,...,10.

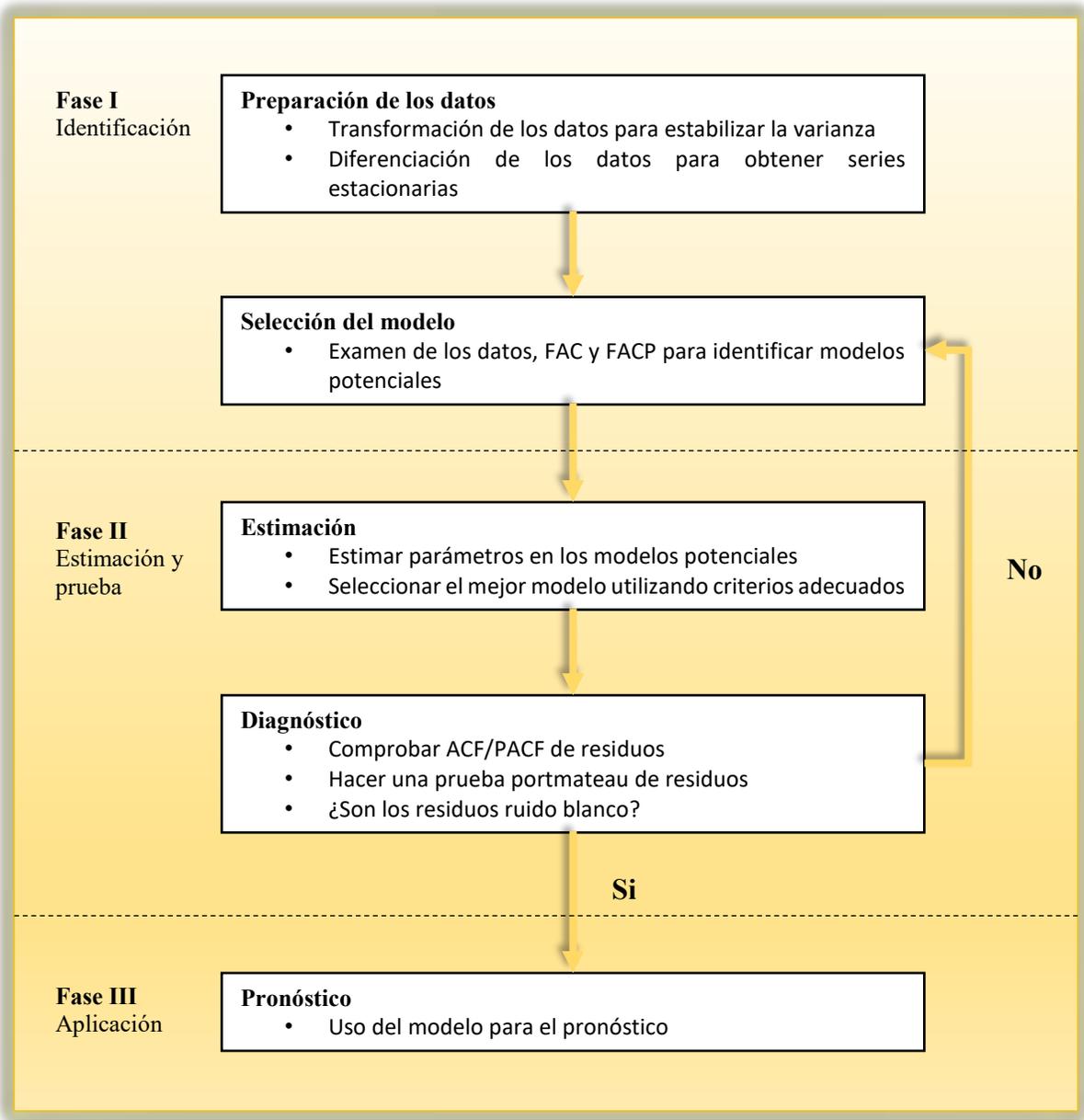


Figura 6. 1. Representación esquemática de la metodología Box-Jenkins para modelado de series de tiempo

Tabla 6. 4. Serie de tiempo con 36 valores

Periodo	Valor	Periodo	Valor	Periodo	Valor
1	23	13	86	25	17
2	59	14	33	26	45
3	36	15	90	27	9

4	99	<b>16</b>	74	<b>28</b>	72
5	36	<b>17</b>	7	<b>29</b>	33
6	74	<b>18</b>	54	<b>30</b>	17
7	30	<b>19</b>	98	<b>31</b>	3
8	54	<b>20</b>	50	<b>32</b>	29
9	17	<b>21</b>	86	<b>33</b>	30
10	36	<b>22</b>	90	<b>34</b>	68
11	89	<b>23</b>	65	<b>35</b>	87
12	77	<b>24</b>	20	<b>36</b>	44

Grafica de serie de Tiempo valores 0-100

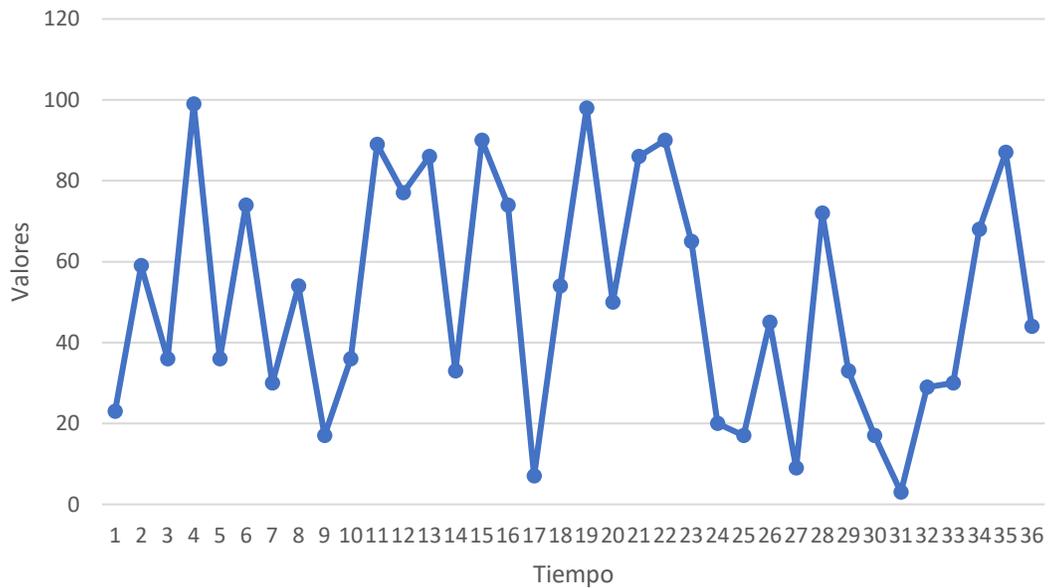


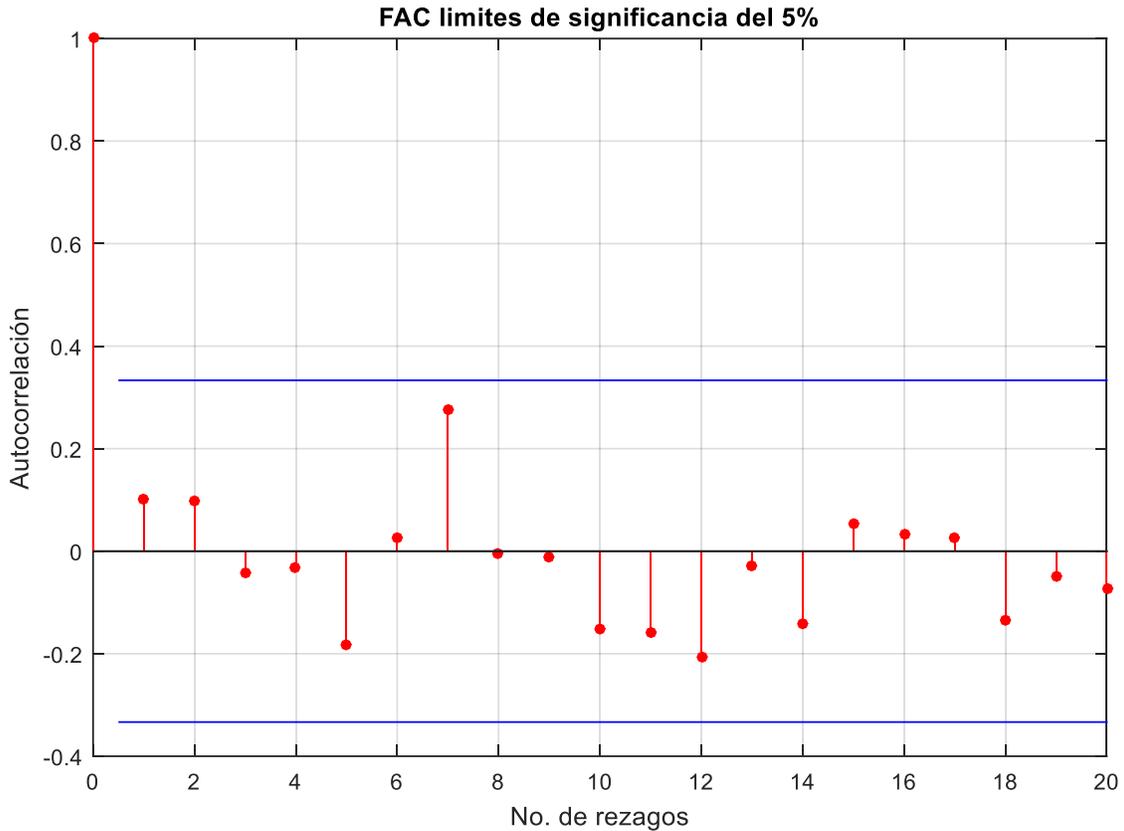
Figura 6.2. Gráfica de tiempo de los datos de la Tabla 6.4

La función de autocorrelación es una herramienta valiosa para investigar las propiedades de una serie de tiempo empírica. Sin embargo, la teoría estadística fundamental de  $r_k$  es bastante complicada, y en algunos casos intratable. Para el caso especial de una serie de “ruido blanco”, la teoría de muestreo de  $r_k$  se conoce y puede ser utilizada para ventajas prácticas [68].

### 6.2.2. Modelo de ruido blanco.

La (Ecuación 6.7) es un modelo simple aleatorio dónde la observación  $Y_t$  se conforman de dos partes, un nivel general,  $c$ , y un componente de error aleatorio,  $e_t$ , que no se correlaciona de periodo a periodo:

$$Y_t = c + e_t \tag{6.7}$$



**Figura 6.3.** Gráfica de la Función de autocorrelación para serie de tiempo de la Tabla 6.4

Los datos de la (Tabla 6.4) se obtienen de este modelo. A menudo se llama modelo de “ruido blanco” – una terminología que viene de ingeniería. El modelo de ruido blanco es fundamental para muchas técnicas en el análisis de series de tiempo. Cualquier modelo de pronóstico bueno debe de tener errores de pronóstico que asemejen un modelo de ruido blanco [68].

**La distribución de la muestra de las autocorrelaciones.**

Con una serie de tiempo que es ruido blanco, la teoría de la muestra de  $r_k$  se conoce y así las propiedades de la FAC pueden ser estudiadas para este modelo.

Una manera de abordar este problema es estudiar los valores de  $r_k$  uno a la vez para desarrollar una fórmula de error estándar para probar si una  $r_k$  particular es significativamente diferente de cero [68].

Teóricamente, todos los coeficientes de autocorrelación para una serie de números aleatorios deben de ser cero. Pero ya que se tienen muestras finitas, cada una de las autocorrelaciones de muestra no será exactamente cero. Ha sido demostrado por Anderson (1942), Bartlett (1946), Quenouille (1949), y otros, que los coeficientes de autocorrelación de datos de ruido blanco tienen una distribución muestral que puede ser aproximada por una curva normal con

media cero y error estándar  $1/\sqrt{n}$  donde  $n$  es el número de observaciones en la serie. Esta información puede ser utilizada para desarrollar pruebas de hipótesis similares a los de las pruebas- $F$  y las pruebas- $T$  [68].

Por ejemplo, 95% de todos los coeficientes de autocorrelación de muestra deben caer entre un rango específico por la media más o menos 1.96 errores estándar. Ya que la media es cero y el error estándar es  $1/\sqrt{n}$  para ruido blanco, se espera que aproximadamente el 95% de todos los coeficientes de autocorrelaciones de la muestra estén entre  $\pm 1.96/\sqrt{n}$ . Si este no es el caso, es probable que la serie no sea ruido blanco. Por esta razón, es común graficar líneas en  $\pm 1.96/\sqrt{n}$  cuando se gráfica la FAC. Estos límites se conocen como valores críticos [68].

En la (Tabla 6.4),  $n = 36$  y así el error estándar es  $1/\sqrt{36} = 0.167$ . Esto quiere decir que los límites para la autocorrelación están en  $\pm 1.96/\sqrt{36} = \pm 0.327$ . Estos límites se muestran en la (Figura 6.3). Los 20 coeficientes de autocorrelación caen entre estos límites, confirmando lo que en este caso ya se sabía los datos son ruido blanco.

El concepto de una distribución muestral es de importancia crítica en el análisis de las series de tiempo. El coeficiente de autocorrelación para un tiempo de retardo de siete periodos en la (Figura 6.3) es de 0.275. Este valor es diferente de cero debido al efecto de la casualidad. La distribución muestral proporciona las directrices sobre lo que es el azar y lo que constituye una relación significativa. El valor de 0.275 no es significativamente diferente de cero. De cualquier modo, si este valor se hubiera obtenido para 360 observaciones en lugar de 36, el error estándar hubiera sido solo de 0.053 y los límites hubieran sido de  $\pm 0.103$ , en lugar de  $\pm 0.327$ . Esto quiere decir que, en promedio, las autocorrelaciones serían más pequeñas que  $\pm 0.103$  en 95 de cada 100 veces. En este caso un  $r_7$  de 0.275 hubiera indicado la presencia de un patrón cada siete retardos (o periodos), ya que hubiera caído fuera de los límites. Desde luego, con 360 valores aleatorios, sería muy poco probable observar un valor de  $r$  tan alto. En resumen, la distribución muestral y el error estándar nos permiten interpretar los resultados del análisis de autocorrelación de tal manera que podemos distinguir lo que es el patrón de lo que es la aleatoriedad, o ruido blanco, en nuestros datos [68].

### Coefficiente de autocorrelación parcial.

En el análisis de regresión, si la variable pronosticada  $Y$  es una regresión de las variables explicativas  $X_1$  y  $X_2$ , entonces pudiera ser de interés preguntar cuanta influencia tiene la variable explicativa  $X_1$  si los efectos de  $X_2$  de alguna manera se eliminarán (*partialled out*) primero. Típicamente, esto quiere decir la regresión de  $Y$  sobre  $X_2$ , obteniendo los errores residuales de este análisis, y encontrando la correlación de los residuos con  $X_1$ . En el análisis de las series de tiempo existe un concepto similar.

Las autocorrelaciones parciales se utilizan para medir el grado de asociación entre  $Y_t$  y  $Y_{t-k}$  cuando los efectos de otros retardos  $-1, 2, 3, \dots, k-1$  son removidos.

El valor de esto se puede ver en el siguiente ejemplo. Suponga que había una autocorrelación significativa entre  $Y_t$  y  $Y_{t-1}$ . Entonces también habrá una correlación significativa entre  $Y_{t-1}$  y  $Y_{t-2}$  ya que estas son también una unidad de tiempo separadas. Consecuentemente, habrá

una correlación entre  $Y_t$  y  $Y_{t-2}$  porque ambos se relacionan a  $Y_{t-1}$ . De tal manera que para medir la autocorrelación real entre  $Y_t$  y  $Y_{t-2}$ , necesitamos eliminar el valor que interviene  $Y_{t-1}$ . Esto es lo que hace la función autocorrelación parcial.

Los coeficientes de autocorrelación parcial de orden  $k$  se denotan por  $\alpha_k$  y se puede calcular mediante la regresión de  $Y_t$  contra  $Y_{t-1}, \dots, Y_{t-k}$ :

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_k Y_{t-k} \quad (6.8)$$

Esta es una regresión inusual debido a que las variables explicativas del lado derecho son valores previos de la variable que se pronostica  $Y_t$ . Estos son simplemente valores rezagados de la variable que se pronostica, y por lo tanto el nombre autorregresión (AR) se utiliza para describir la ecuación de la forma (6.8) [68].

La autocorrelación parcial,  $\alpha_k$ , es un coeficiente estimado  $b_k$  de esta regresión múltiple. Nos damos cuenta de que la primera autocorrelación parcial es siempre igual a la primera autocorrelación. Variando el número de términos del lado derecho de (6.8) nos dará las autocorrelaciones parciales para diferentes valores de  $k$ .

Es útil graficar la función de autocorrelación parcial o FACP. La función de autocorrelación parcial de los datos de la (Tabla 6.4) se gráfica en la Figura 6.4.

Como con la FAC, las autocorrelaciones parciales todas deberían ser aproximadamente cero para las series de ruido blanco. Quenouille (1949) demostró que, si la serie de tiempo es ruido blanco, entonces las autocorrelaciones parciales estimadas son aproximadamente independientes y normalmente distribuidas con un error estándar  $1/\sqrt{n}$ . Por lo tanto, los mismos valores críticos de  $\pm 1.96/\sqrt{n}$  pueden ser utilizadas con una FACP para evaluar si los datos son ruido blanco.

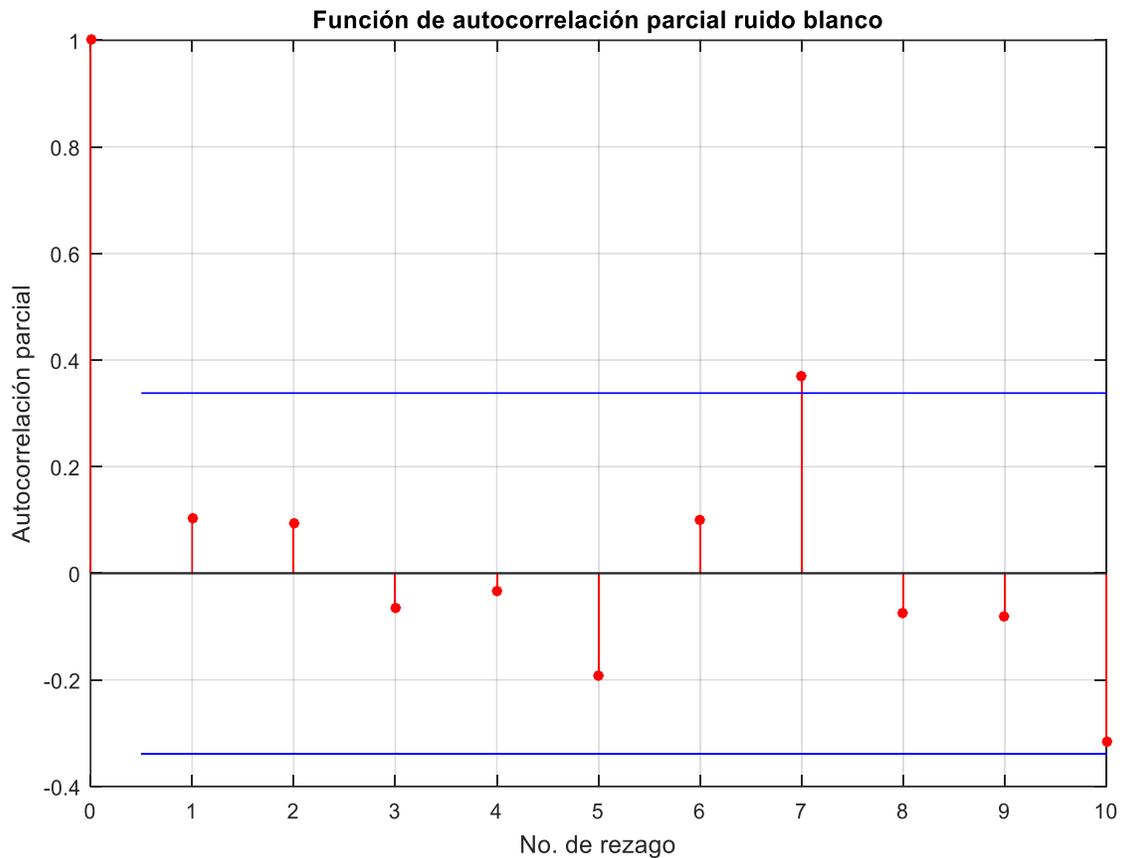
### 6.2.3. Identificación de la estacionalidad en las series de tiempo.

La estacionalidad se define como un patrón que se repite así mismo sobre intervalos fijos de tiempo. Las ventas de combustible para calentadores, por ejemplo, son altas en invierno y bajas en verano, indicando un patrón estacional de 12 meses. Si el patrón es una constante, el coeficiente de autocorrelación en el retraso de 12 meses tendrá un valor positivo alto indicando la existencia de estacionalidad. Si no fuera significativamente diferente de cero, indicarían que los meses con un año de diferencia no están correlacionados con ningún patrón emergente existente de un año al siguiente. Tales datos no serán estacionales.

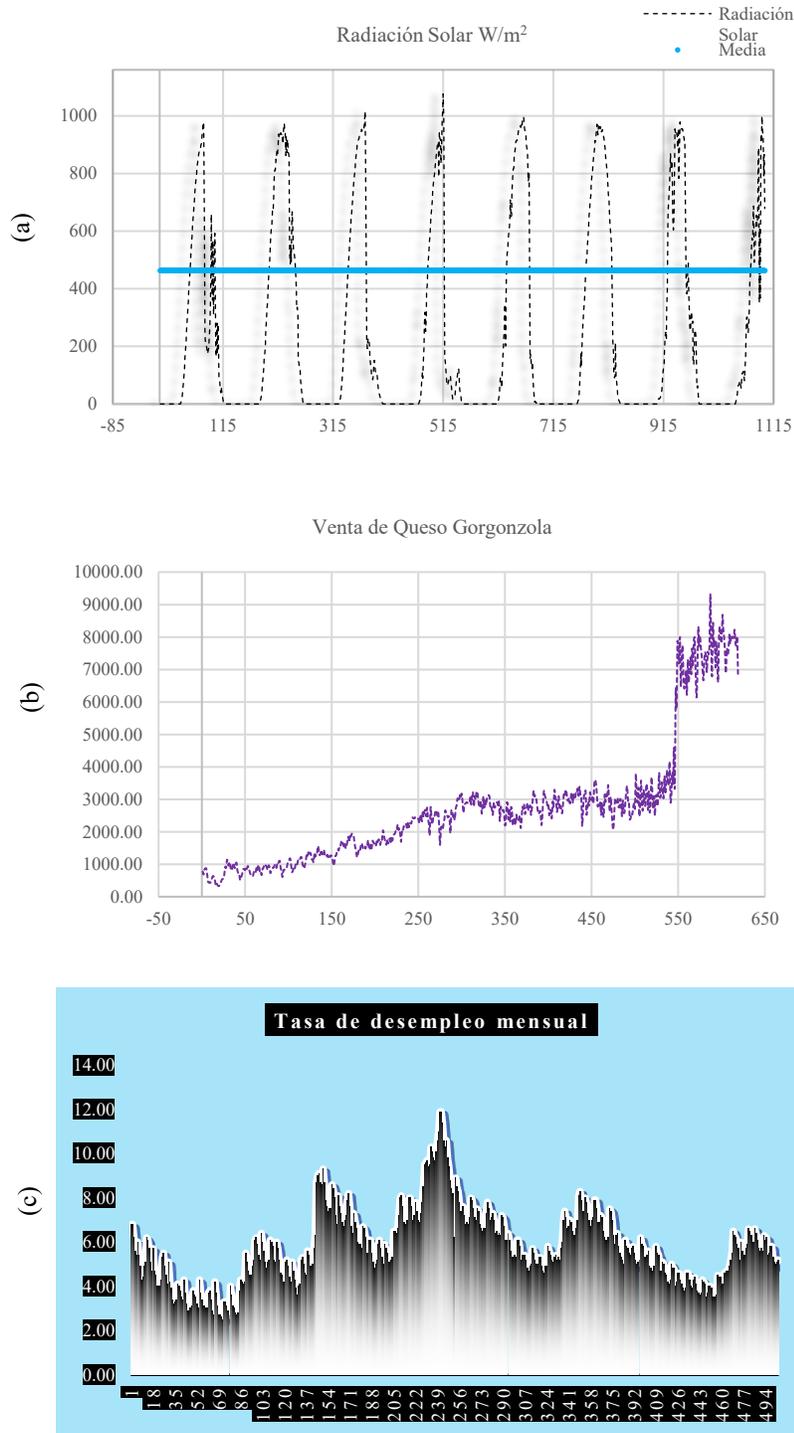
En general, se puede encontrar estacionalidad al identificar un coeficiente de autocorrelación grande o un coeficiente de autocorrelación parcial grande en el desfase estacional. A menudo las autocorrelaciones en múltiples desfases estacionales serán también significativas. Así para los datos mensuales, autocorrelaciones altas pudieran también encontrarse en el desfase de 24 e incluso en el de 36 [69].

## Evaluación de la estacionariedad en los datos de series de tiempo

Recordando que estacionariedad quiere decir que no existe crecimiento o disminución en los datos. Los datos deberían ser aproximadamente horizontales a lo largo del eje de tiempo. En otras palabras, los datos fluctúan alrededor de una media constante, independiente del tiempo, y la varianza de la fluctuación permanece esencialmente constante a lo largo del tiempo. Para una definición formal de estacionariedad, véase Box-Jenkins y Reinsel 1994 [68].



**Figura 6. 4.** Autocorrelaciones parciales para la serie de la Tabla 5.4



**Figura 6. 5.** Datos de series de tiempo; (a) serie de tiempo estacionaria en la media; (b) serie de tiempo no estacionaria en la media; y (c) serie de tiempo no estacionaria en la media ni en la varianza

Normalmente podemos evaluar la estacionariedad utilizando una gráfica de tiempo.

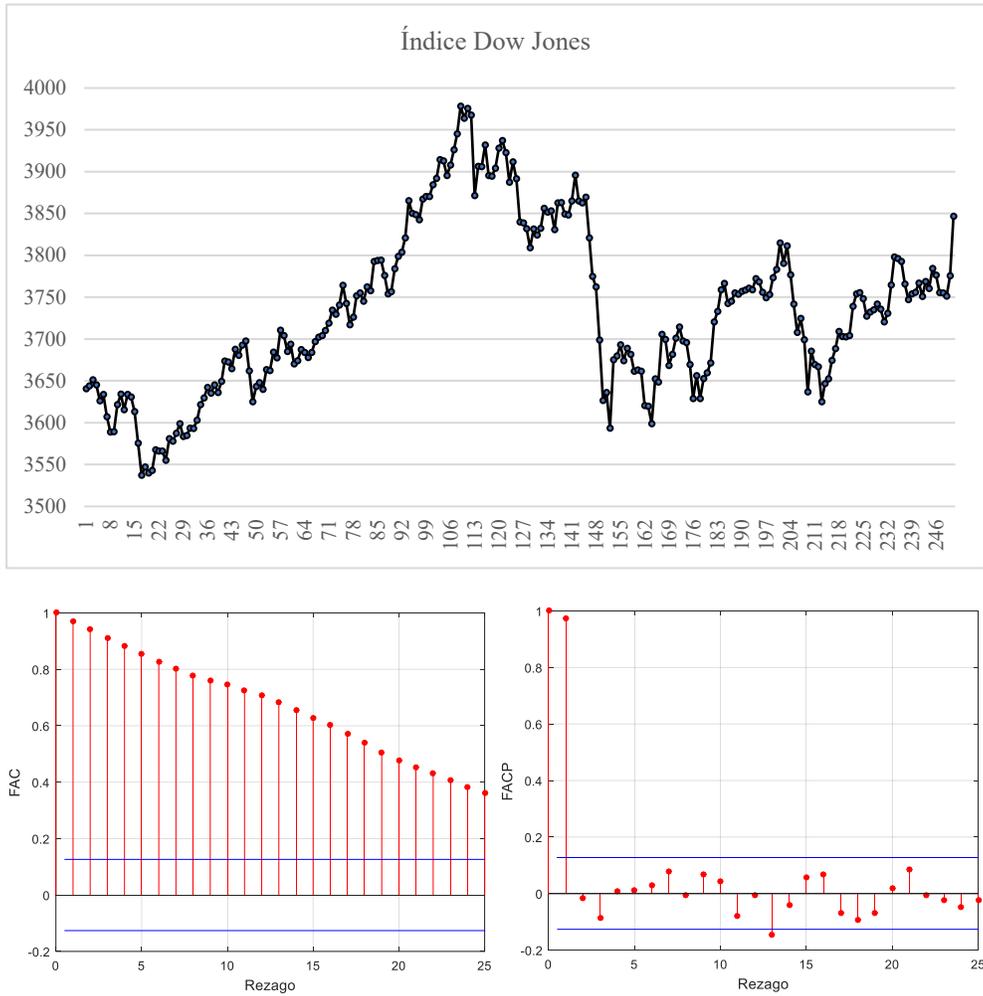
1. Si una serie de tiempo se gráfica y no hay evidencia de un cambio en la media a lo largo del tiempo (**Figura 6. 5a**), se dice que la serie es estacionaria en la media.

2. Si la serie graficada muestra un cambio poco evidente en la varianza a lo largo del tiempo, se dice que la serie es estacionaria en la varianza.

La **Figura 6. 5b** muestra una típica serie de tiempo que no es estacionaria en la media y la media cambia a lo largo del tiempo. En la **Figura 6. 5c** se muestra una serie de tiempo que es no estacionaria en ambas la media y la varianza. La media varia (cambia a lo largo del tiempo), y la varianza (o desviación estándar) no es razonablemente constante a lo largo del tiempo [68].

La grafica de la serie de tiempo normalmente es suficiente para convencer a un pronosticador de que los datos son estacionarios o no estacionarios. La gráfica de autocorrelación puede incluso demostrar fácilmente la no estacionariedad en la media. Las autocorrelaciones de datos estacionarios caen hasta cero relativamente rápido, mientras que las series no estacionarias son significativamente diferentes de cero para muchos desfases de tiempo. Cuando se representan gráficamente, las autocorrelaciones de datos no estacionarios decrecen lentamente según aumente el número de desfases de tiempo [68].

La **Figura 6. 6** muestra el índice Dow-Jones a lo largo de 251 días de actividad. La gráfica de tiempo muestra que es no estacionario en la media. La FAC también muestra un patrón típico para una serie no estacionaria, con una lenta disminución en el tamaño de la autocorrelación. La autocorrelación para un desfase de tiempo,  $r_1$ , es muy grande y positivo. La autocorrelación para dos desfases de tiempo también es grande y positiva, pero no tan grande como para  $r_1$ , porque la componente del error aleatorio ha entrado dos veces en el cuadro. En general,  $r_k$  para datos no estacionarios será relativamente grande y positivo, hasta que  $k$  sea suficientemente grande de manera que el error aleatorio comience a dominar la autocorrelación. La FACP presentada en la **Figura 6. 6** también es típica de una serie no estacionaria con un pico alto cercano a uno en el desfase 1 [68].



**Figura 6. 6.** Índice Dow Jones para 251 días de mercado finalizando el 26 de agosto de 1994

### 6.2.4. Remover la no estacionariedad en una serie de tiempo

Tendencias, u otros patrones no estacionarios en el nivel de una serie, resulta en autocorrelaciones positivas que dominan el diagrama de autocorrelación. Por consiguiente, es importante eliminar la no-estacionariedad, de tal manera que otras estructuras de autocorrelación puedan ser vistas antes de proseguir con la construcción del modelo de la serie de tiempo. Una manera de eliminar la no-estacionariedad es a través del método de diferenciación. La serie diferenciada se define como el cambio entre cada observación en la serie original [68].

$$Y'_t = Y_t - Y_{t-1} \tag{6.9}$$

La serie diferenciada únicamente tendrá  $n - 1$  valores ya que no es posible calcular una diferencia  $Y_1'$  para la primera observación [68].

Tomando las primeras diferencias es una herramienta muy útil para remover la no-estacionariedad. Sin embargo, ocasionalmente los datos diferenciados no aparecerán estacionarios y puede ser necesario diferenciar los datos una segunda vez [68].

$$Y_1'' = Y_1' - Y_{t-1}' = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}.$$

$Y_t''$  se refiere a como las series de diferencias de segundo orden. Estas series tendrán  $n - 2$  valores. En la práctica, casi nunca se necesita ir más allá de las diferencias de segundo orden, porque los datos reales generalmente involucran la no estacionariedad únicamente del primero o segundo nivel [68].

### 6.2.5. Diferenciación estacional

Con datos estacionales que no son estacionarios, puede ser apropiado tomar diferencias estacionales. Una diferencia estacional es la diferencia entre una observación y la observación correspondiente del año anterior. Así que para datos mensuales teniendo un patrón anual de 12 meses, obtenemos

$$Y_t' = Y_t - Y_{t-12} \quad (6.10)$$

En general, las series diferenciales estacionarias,  $Y_t'$ , es el cambio entre observaciones separadas por  $s$  periodos de tiempo, donde  $s$  es el número de estaciones. Para datos mensuales  $s = 12$ , para datos trimestrales  $s = 4$ , y así sucesivamente. Al igual que con las primeras diferencias, la diferenciación puede ser repetida para obtener una diferenciación estacional de segundo-orden, aunque esto es raramente requerido [68].

### 6.2.6. Notación de desplazamiento hacia atrás “Backshift”

Un dispositivo de notación muy útil es el operador de desplazamiento hacia atrás,  $B$ , que se utiliza como sigue:

$$BY_t = Y_{t-1} \quad (6.11)$$

En otras palabras,  $B$ , operando en  $Y_t$ , tiene el efecto de desplazar los datos un periodo atrás. Aplicar dos veces  $B$  a  $Y_t$  desplaza los datos dos periodos atrás, como sigue:

$$B(BY_t) = B^2Y_t = Y_{t-2} \quad (6.12)$$

Para datos mensuales, si queremos cambiar la atención a “el mismo mes del año pasado” entonces  $B^{12}Y_t = Y_{t-12}$ .

El operador de desplazamiento hacia atrás es conveniente para describir el proceso de diferenciación. Una primera diferencia puede ser escrita como

$$Y'_t = Y_t - Y_{t-1} = Y_t - BY_t = (1 - B)Y_t \quad (6.13)$$

Nos damos cuenta de que una primera diferencia se representa por  $(1 - B)$ . De manera similar, si hacemos una diferencia de segundo-orden (i.e, la primera diferencia de la primera diferencia) tiene que ser calculada, entonces:

$$\begin{aligned} Y''_t &= (Y'_t - Y'_{t-1}) \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2} \\ &= (1 - 2B + B^2)Y_t \\ Y''_t &= (Y'_t - Y'_{t-1}) \\ &= (1 - B)^2Y_t \end{aligned} \quad (6.14)$$

Nos podemos dar cuenta que la diferencia de segundo orden se denota  $(1 - B)^2$ . (Es importante reconocer que la diferencia de segundo orden no es lo mismo que una segunda diferencia, la cual sería denotada como  $1 - B^2$ ; de manera similar, una duodécima diferencia sería  $1 - B^{12}$ , pero una diferencia de duodécimo orden sería  $(1 - B)^{12}$ ).

En general, una diferencia de d-ésimo orden puede ser escrita como

$$(1 - B)^d Y_t \quad (6.15)$$

Una diferencia estacional seguida por una primera diferencia puede ser escrita como

$$(1 - B)(1 - B^s)Y_t \quad (6.16)$$

La notación “backshift” es conveniente porque los términos pueden ser multiplicados juntos para ver el efecto combinado. Por ejemplo (6.16) puede ser expandida para dar [68].

$$(1 - B)(1 - B^s)Y_t = (1 - B - B^s + B^{s+1})Y_t = Y_t - Y_{t-1} - Y_{t-s} + Y_{t-s-1} \quad (6.17)$$

### 6.2.7. Modelos ARIMA para datos de series de tiempo

La (Ecuación 6.18) muestra el modelo de regresión de la forma

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p + e \quad (6.18)$$

Dónde  $Y$  es la variable a pronosticar, de  $X_1$  hasta  $X_p$ , son las variables explicativas, de  $b_0$  hasta  $b_p$  son los coeficientes de regresión lineal, y  $e$  es el término de error. En la ([Ecuación 6.18](#)) las variables  $X$  pueden representar cualquier factor como GNP, precios publicitarios, activos, etc [68].

Suponga, sin embargo, que esas variables están definidas como  $X_1 = Y_{t-1}$ ,  $X_2 = Y_{t-2}$ ,  $X_3 = Y_{t-3}$ , ...,  $X_p = Y_{t-p}$  entonces la ([Ecuación 6.18](#)) se convierte en la ([Ecuación 6.19](#))

$$Y_t = b_0 + b_1Y_{t-1} + b_2Y_{t-2} + \dots + b_pY_{t-p} + e_t \quad (6.19)$$

La ([Ecuación 6.19](#)) sigue siendo una ecuación de regresión, pero difiere de ([Ecuación 6.18](#)) en que el lado derecho de la ([Ecuación 6.18](#)) son diferentes variables explicativas, mientras que las de ([6.18](#)) son valores previos de la variable a pronosticar. Éstos son simplemente valores desfasados de la variable a pronosticar y por lo tanto se utiliza autorregresión (AR) para referirse a ecuaciones de la forma de ([6.19](#)) [68].

Una pregunta que surge por considerar la ([Ecuación 6.19](#)) es porque la autorregresión se debe tratar diferente a una regresión ordinaria. La respuesta es:

1. En la autorregresión la suposición básica de independencia de los términos de error (residual) pueden ser fácilmente violados dado que las variables explicativas en la ([Ecuación 6.19](#)) normalmente tienen una relación de dependencia incorporada.
2. Determinar el número de valores pasados de  $Y_t$  para incluirlos en la ([Ecuación 6.19](#)) no siempre es fácil.

Así como es posible la regresión contra valores pasados de la serie, hay un modelo de serie de tiempo que utiliza valores pasados como variables explicativas:

$$Y_t = b_0 + b_1e_{t-1} + b_2e_{t-2} + \dots + b_qe_{t-q} + e_t \quad (6.20)$$

Aquí, de manera explícita, una relación de dependencia se establece entre los términos de error sucesivos, y la ecuación es llamada un modelo de promedio móvil.

El modelo en ([6.20](#)) es conocido como un promedio móvil porque se define como el promedio móvil del modelo del error de la serie,  $e_t$  [68].

Los modelos autorregresivos (AR) se pueden acoplar de manera efectiva con los modelos de promedios móviles (MA) para formar una clase general y útil de modelos de series de tiempo conocidos como modelos autorregresivos de promedios móviles (ARMA). De cualquier forma, estos pueden ser utilizados únicamente cuando los datos son estacionarios. Esta clase de modelos pueden ser extendidas a series no estacionarias por lo que permite la diferenciación de los datos de la serie. Estos se conocen como modelos autorregresivos integrados de promedios móviles (ARIMA) [68].

Existe una gran variedad de modelos ARIMA. El modelo general no estacional se conoce como ARIMA  $(p, d, q)$ :

- AR:  $p$  = orden de la parte autorregresiva  
 I:  $d$  = grado de la primera diferencia involucrada  
 MA:  $q$  = orden de la parte de promedio móvil

Los dos modelos que ya se han visto pueden ser escritos en esta notación. Un modelo de ruido blanco se clasifica como ARIMA  $(0,0,0)$  porque no hay ninguna parte AR a este ( $Y_t$  no depende de  $Y_{t-1}$ ), no existe ninguna diferencia involucrada, y no existe ninguna parte MA ( $Y_t$  no depende de  $e_{t-1}$ ). De manera similar, un modelo de caminata aleatoria se clasifica como ARIMA  $(0,1,0)$  porque no tiene parte AR o MA e involucra una diferencia [68].

Nos damos cuenta de que si  $p$ ,  $d$  o  $q$  son iguales a cero, el modelo puede también ser escrito de forma abreviada al eliminar los parámetros que no son necesarios del modelo. Por ejemplo un ARIMA $(2,0,0)$  puede ser escrito como AR(2) porque no existe diferenciación (la parte I) y la parte de promedio móvil (MA). De forma similar, un ARIMA  $(1,0,1)$  puede también ser escrita como ARMA $(1,1)$ ; un ARIMA  $(0,1,1)$  puede ser escrito también como IMA $(1,1)$  y así sucesivamente [68].

### Modelo autorregresivo de orden uno

La (Ecuación 6.20) muestra la forma básica de un modelo ARIMA $(1,0,0)$  o AR(1). La observación  $Y_t$  depende de  $Y_{t-1}$ , y el valor de coeficiente autorregresivo  $\phi_1$  está restringido para caer entre  $-1$  y  $+1$ :

$$Y_t = c + \phi_1 Y_{t-1} + e_t \quad (6.20)$$

La gráfica de tiempo de un modelo AR(1) varía con el parámetro  $\phi_1$ . Por ejemplo, cuando  $\phi_1 = 0$ ,  $Y_t$  es equivalente a una serie de ruido blanco (comparar la (Ecuación 6.7)). Cuando  $\phi_1 = 1$ ,  $Y_t$  es equivalente a una serie de caminata aleatoria [68].

### Modelo de promedio móvil de orden uno

La Ecuación 6.21 da un modelo MA(1) o ARIMA $(0,0,1)$  para ser general. La observación  $Y_t$  depende del término de error  $e_t$  y también el término de error anterior  $e_{t-1}$  con coeficiente  $-\theta_1$ :

$$Y_t = c + e_t - \theta_1 e_{t-1} \quad (6.21)$$

El valor del coeficiente  $\theta_1$  está restringido a caer entre  $-1$  y  $+1$ . Nos damos cuenta del signo menos en el coeficiente  $\theta_1$  en (6.21). Esto es una convención para los modelos ARIMA.

$$Y_t = 10 + e_t - 0.7e_{t-1} \tag{6.22}$$

Donde  $e_t$  esta normalmente distribuida con media cero y varianza 1. En este caso para el modelo MA (1) (Ecuación 5.22) tenemos que  $\theta_1 = 0.7$  [68].

### Modelos autorregresivos de orden superior

De manera general, un  $p$  – ésimo modelo AR se define como sigue:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \tag{6.23}$$

Dónde  $c$  = termino constante, ARIMA( $p, 0, 0$ ) o AR( $p$ )  
 $\phi_j = j$  – ésimo parámetro auto-regresivo  
 $e_t$  = el termino de error en el tiempo  $t$ .

Existen restricciones específicas sobre los valores permisibles de los parámetros auto-regresivos. Para  $p = 1$ ,  $-1 < \phi_1 < 1$ . Para  $p = 2$ , las siguientes tres condiciones deben ser cumplidas:

$$-1 < \phi_2 < 1 \qquad \phi_2 + \phi_1 < 1 \qquad \phi_2 - \phi_1 < 1.$$

Condiciones más complicadas se sostienen para  $p \geq 3$ . Utilizando el símbolo de desplazamiento hacia atrás,  $B$ , (Ecuación 6.23) puede reescribirse como

$$\left. \begin{aligned} & Y_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} = c + e_t \\ \text{o} & (1 - \phi_1 B - \dots - \phi_p B^p) Y_t = c + e_t \end{aligned} \right\} \tag{6.24}$$

Una gran variedad de series de tiempo es posible con los modelos autoregresivos. La (Ecuación 6.25) muestra un ejemplo, un modelo AR (2) con la siguiente ecuación:

$$Y_t = 6 + 1.2Y_{t-1} - 0.8Y_{t-2} + e_t \tag{6.25}$$

Donde  $e_t$  fue generada de una distribución normal con media cero y varianza 1. Nos damos cuenta que para el modelo AR (2), cae en un modo senoidal amortiguado y que hay exactamente dos correlaciones parciales significativas [68].

En general, la FAC de un modelo AR ( $p$ ) con  $p \geq 2$  puede mostrar un decaimiento exponencial o patrones senoidales. El modelo de auto – correlaciones de un modelo AR ( $p$ ) son cero más allá del retardo  $p$  [68].

**Tabla 6. 5.** Patrones esperados en la FAC y la FACP para modelos AR y MA individuales

Proceso	FAC	FACP
<b>AR (1)</b>	Decaimiento exponencial: sobre el lado positivo si $\phi_1 > 0$ y alternando de signo a partir del lado negativo si $\phi_1 < 0$ .	Espiga en el retardo 1, luego se reduce a cero: espiga positiva si $\phi_1 > 0$ , negativo si $\phi_1 < 0$ .
<b>AR (p)</b>	Decaimiento exponencial o sinusoidal amortiguada. El patrón exacto depende de los signos y dimensiones y tamaños de $\phi_1, \dots, \phi_p$ .	Espigas en el retardo 1 hasta $p$ , luego se reduce hasta cero.
<b>MA (1)</b>	Espiga en el retardo 1 entonces se reduce a cero: espiga positiva si $\theta_1 < 0$ , negativo si $\theta_1 > 0$ .	Decaimiento exponencial: sobre el lado negativo si $\theta_1 > 0$ y alternancia de signo a partir del lado positivo si $\theta_1 < 0$ .
<b>MA (q)</b>	Espigas en retardo 1 hasta $q$ , entonces se reduce a cero.	Decaimiento exponencial o sinusoidal amortiguado. El patrón exacto depende de los signos y el tamaño de $\theta_1 \dots \theta_p$ .

### Modelos de promedios móviles de orden superior

El modelo general MA de orden  $q$  puede ser escrito como sigue:

$$Y_t = c + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \tag{6.26}$$

Donde  $c$  = termino constante,  
 $\theta_j = j$  – Ésimo parámetro de promedio móvil,  
 $e_{t-k}$  = el termino de error en el tiempo  $t - k$ .

Las mismas restricciones que fueron requeridas para los modelos AR son también requeridas para los modelos MA. Por lo tanto, para  $q = 1$ , nosotros requerimos  $-1 < \theta_1 < 1$ . Para  $q = 2$ , las siguientes consideraciones todas deben cumplirse:

$$-1 < \theta_2 < 1 \quad \theta_2 + \theta_1 < 1 \quad \theta_2 - \theta_1 < 1.$$

Condiciones más complicadas se mantienen para  $q \geq 3$ .  
 En notación de desplazamiento hacia atrás, (6.26) puede ser escrita

$$Y = c + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) e_t.$$

Una amplia variedad de series de tiempo puede ser producidas utilizando modelos de promedios móviles. La (Ecuación 6.27) muestra un ejemplo de un modelo de promedios móviles [68].

$$Y_t = e_t - e_{t-1} + 0.8e_{t-2} \quad (6.27)$$

Nos damos cuenta de que existen dos autocorrelaciones no cero y que las autocorrelaciones parciales decaen en forma sinusoidal. Esto es exactamente lo contrario de lo que ocurre con un modelo AR (2). En general, las autocorrelaciones de un modelo MA( $q$ ) son cero más allá del retardo  $q$ . Si  $q \geq 2$ , la FACP puede mostrar decaimiento exponencial o patrones de amortiguamiento senoidal [68].

### Combinación: modelos ARMA

Los elementos básicos de los modelos AR y MA pueden ser combinados para producir una gran variedad de modelos. Por ejemplo, la (Ecuación 6.28) combina los modelos AR de primer orden y un modelo MA de primer orden. Esto se conoce como modelo ARMA (1,1) o ARIMA (1,0,1):

$$Y_t = c + \phi_1 Y_{t-1} + e_t - \theta_1 e_{t-1} \quad (6.28)$$

Aquí,  $Y_t$  depende de un valor previo  $Y_{t-1}$  y un término de error previo  $e_{t-1}$ . La serie se asume como estacionaria en la media y en la varianza [68].

Utilizando la notación de desplazamiento hacia atrás, el ARIMA (1,0,1) se escribe

$$(1 - \phi_1 B) Y_t = c + (1 - \theta_1 B) e_t$$

$\uparrow$   
AR(1)

$\uparrow$   
MA(1)

Un modelo ARMA con términos de orden superior se escribe

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad \text{ARIMA } (p, 0, q)$$

o

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) e_t \quad (6.29)$$

Los mismos parámetros de restricción se aplican aquí como para los modelos AR y MA puros [68].

### Combinación: Modelos ARIMA

Si se agrega no estacionariedad a un modelo combinado ARMA, entonces el modelo general ARIMA  $(p, i, q)$  es obtenido. La ecuación para el caso más simple, ARIMA(1,1,1) es como sigue:

$$(1 - \phi_1 B)(1 - B)Y_t = c + (1 - \theta_1 B)e_t$$

$\uparrow$   
AR(1)

$\uparrow$   
Primera  
diferencia

$\uparrow$   
MA(1)

Dese cuenta de que se usa el operador de desplazamiento hacia atrás para describir (i) la primera diferencia, (ii) la porción del modelo AR(1), y (iii) el aspecto MA(1) [68].

El modelo general ARIMA  $(p, d, q)$  produce una tremenda variedad de patrones en las FAC y FACP, así que no es sabio establecer reglas para identificar modelos ARIMA generales. De cualquier modo, los modelos más simples AR( $p$ ) y MA ( $q$ ) proporcionan algunas características que pueden ayudar a un pronosticador novato en una identificación del modelo ARIMA. Esto también es útil para saber que varios modelos diferentes pueden producir casi la misma cantidad de pronósticos, de tal manera que el proceso de identificación no es como buscar una aguja en un pajar [68].

En la práctica, raramente es necesario tratar con valores de  $p, d,$  o  $q$  que sean otros que 0,1, o 2. Quizás sea notable que como un pequeño rango de valores para  $p, d,$  o  $q$  pueda cubrir un tremendo rango de situaciones prácticas de pronóstico [68].

### Estacionalidad y modelos ARIMA

Una última complejidad que agregar a los modelos ARIMA es la estacionalidad. Exactamente de la misma manera que los puntos de datos consecutivos pueden exhibir propiedades AR, MA, combinación ARMA, o combinación ARIMA, de tal manera, que los datos separados por una estación entera (i.e. un año) pudieran exhibir las mismas propiedades [68].

La notación ARIMA puede ser extendida fácilmente para manejar aspectos estacionales, y la notación general abreviada es

$$\text{ARIMA}(p, d, q) (P, D, Q)_s$$

$\uparrow$   
(
)  
 Parte del  
modelo no  
estacional

$\swarrow$   
(
)  
 Parte del  
modelo  
estacional

Dónde  $s =$  es el número de periodos por estación [68].

El álgebra es simple, pero puede llegar a extenderse, de tal manera que para propósitos ilustrativos considere el siguiente modelo general ARIMA  $(1,1,1)(1,1,1)_4$ :

$$\begin{array}{c}
 \left( \begin{array}{c} \text{Diferencia} \\ \text{No-estacional} \end{array} \right) \qquad \qquad \qquad \left( \begin{array}{c} \text{MA(1)} \\ \text{estacional} \end{array} \right) \\
 \downarrow \qquad \qquad \qquad \qquad \qquad \qquad \downarrow \\
 (1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)Y_t = (1 - \theta_1 B)(1 - \Theta_1 B^4)e_t \\
 \uparrow \qquad \qquad \qquad \uparrow \qquad \qquad \qquad \uparrow \qquad \qquad \qquad \uparrow \\
 \left( \begin{array}{c} \text{AR (1)} \\ \text{No-estacional} \end{array} \right) \left( \begin{array}{c} \text{Estacional} \\ \text{AR(1)} \end{array} \right) \left( \begin{array}{c} \text{Diferencia} \\ \text{estacional} \end{array} \right) \left( \begin{array}{c} \text{MA(1)} \\ \text{No-estacional} \end{array} \right)
 \end{array}$$

Todos los factores pueden ser multiplicados y el modelo general escrito como sigue:

$$\begin{aligned}
 Y_t = & (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + (1 + \Phi_1)Y_{t-4} \\
 & - (1 + \phi_1 + \Phi_1 + \phi_1 \Phi_1)Y_{t-5} + (\phi_1 + \phi_1 \Phi_1)Y_{t-6} \\
 & - \Phi_1 Y_{t-8} + (\Phi_1 + \phi_1 \Phi_1)Y_{t-9} - \phi_1 \Phi_1 Y_{t-10} + e_t - \theta_1 e_{t-1} \\
 & - \Theta_1 e_{t-4} + \theta_1 \Theta_1 e_{t-5}
 \end{aligned} \tag{6.30}$$

De esta manera, una vez que los coeficientes  $\phi_1, \Phi_1, \theta_1$  y  $\Theta_1$  han sido estimados de los datos, la (Ecuación 6.30) puede ser utilizada para el pronóstico.

Nos damos cuenta de que el término constante ha sido omitido para claridad. Si  $Y_t$  es remplazado por  $(Y_t - \mu)$ , donde  $\mu$  es la media de los valores  $Y$ , entonces un término constante aparecerá en última instancia sobre el lado derecho de la (Ecuación 6.30).

La parte estacional de un modelo AR o MA será visto en los retardos estacionales de la FACP y de la FAC. Por ejemplo, el modelo estacional MA, ARIMA  $(0,0,0)(1,0,0)_{12}$  mostrará una espiga en el retardo 12 en la FAC pero no otras espigas significativas. La FACP mostrará decaimiento exponencial en los retardos estacionales; esto es, en los retardos 12, 24, 36,.... De forma similar un ARIMA  $(0,0,0)(1,0,0)_{12}$  (un modelo AR estacional) mostrará decaimiento exponencial en los retardos estacionales de la FAC, y un solo pico significativo en el rezago 12 en la FACP [68].

### 6.3 Pruebas de desempeño

Ya que las técnicas cuantitativas de pronóstico implican, por lo regular, series de tiempo de datos, se desarrolló una notación matemática para hacer referencia a cada periodo específico.

#### Notación básica del pronóstico

$Y_t =$  Valor de una serie de tiempo en el periodo  $t$

$\hat{Y}_t =$  Valor del pronóstico para  $Y_t$

$e_t = \hat{Y}_t - Y_t =$  residual o error de pronóstico

### 6.3.1 Error medio (ME)

Este valor por lo general es muy cercano a cero, debido a que los valores positivos y negativos tienden a eliminarse entre sí. Por otro lado, un valor cercano a cero puede indicar que el pronóstico es adecuado, ya que los errores de pronóstico deben comportarse como ruido blanco. La [ecuación 6.31](#) se utiliza para calcular el ME [67][68].

$$\text{ME} = \frac{1}{n} \sum_{t=1}^n e_t \quad (6.31)$$

### 6.3.2 Error medio absoluto (MAE)

El error absoluto se define, primero, haciendo positivos los valores y después promediando los resultados. La [ecuación 6.32](#) se emplea con este propósito [67][68].

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (6.32)$$

### 6.3.3 Error cuadrado medio (MSE)

Esta prueba se lleva a cabo elevando cada uno los errores al cuadrado, al igual que la prueba anterior, con el propósito de que los errores no se eliminen entre sí. La única diferencia es que esta prueba penaliza los errores grandes. La [ecuación 6.33](#) se utiliza para calcular el MSE [67][68].

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad (6.33)$$

### 6.3.4 Raíz del error cuadrado medio (RMSE)

Se obtiene al sacar la raíz del MSE y es un indicativo de que tan dispersos se encuentra los datos de la línea de regresión, así como de la efectividad del modelo al momento de ser utilizado en pronósticos a corto plazo. La [ecuación 6.34](#) se utiliza con este propósito [70].

$$\text{RMSE} = \frac{1}{n} \sum_{t=1}^n \sqrt{e_t^2} = \sqrt{\text{MSE}} \quad (6.34)$$

# Capítulo

## 7

### Redes neuronales artificiales

#### 7.1 Una introducción a la computación neuronal

El principal propósito de todos los sistemas es el control centralizado de varias funciones biológicas, algunas de ellas responsables del abastecimiento de energía, pues el sistema neuronal está conectado al metabolismo, el control cardiovascular y la respiración. Hay mecanismos neuronales para los ritmos biológicos, funciones de emergencia, etc [71].

Las funciones anteriores son comunes en la mayoría de los animales, y en las estructuras biológicas neuronales es posible incluso discernir varios sedimentos de las diferentes fases de la evaluación. Por otro lado, en los animales superiores, la mayor capacidad del sistema nervioso central está relacionada con el comportamiento; esto es, control del estado del organismo con respecto a su entorno, el cual implica muchas tareas diferentes gracias al encadenamiento de acciones elementales para conseguir un comportamiento social complicado [71].

No obstante, cuando se habla de computación neuronal, normalmente sólo se tienen en mente las funciones sensoriales y motoras, así como algún tipo de proceso interno llamado pensamiento. Todas estas funciones son interdependientes en un sentido u otro, puede ser posible conceptualizar alguna de ellas de una forma idealizada [71].

En el desarrollo de la tecnología de la información, actualmente parece existir una nueva fase donde el objetivo es reproducir muchas de estas funciones neuronales artificialmente. No obstante, no está siempre claro a cuáles de los aspectos arriba indicados se refiere. El motivo principal parece ser el desarrollo de nuevos tipos de ordenadores. Por ejemplo, un objetivo puede ser la implementación de funciones sensoriales artificiales para hacer ver u oír a las máquinas; ésta es una extensión de las más tradicionales técnicas de instrumentación [71].

Aunque el potencial de los métodos de inteligencia artificial, por medio de los cuales estos problemas fueron abordados, han sido conocidos desde hace veinticinco años, todavía se espera que nuevas vías para la inteligencia artificial se puedan abrir cuando el paralelismo masivo de los circuitos computacionales y las nuevas tecnologías sean desarrolladas y la capacidad de computacionales y las nuevas tecnologías sean desarrolladas y la capacidad de computación se incremente en ordenes de magnitud. Sin embargo, antes de esto, será necesario encontrar qué se quiere computar. Hay al menos una nueva dimensión de computación que ha sido muy difícil de alcanzar con los ordenadores digitales [72].

#### 7.2 Definición de una red neuronal

Existen numerosas formas de definir lo que son las redes neuronales, desde las definiciones cortas y genéricas hasta las que intentan explicar más detalladamente lo que significa red neuronal o computación neuronal. Veamos algunos ejemplos de ambos casos:

Una nueva forma de computación, inspirada en modelos biológicos.

Un modelo matemático compuesto por un gran número de elementos procesales organizados en niveles.

“...un sistema de computación hecho por un gran número de elementos simples, elementos de proceso muy interconectados, los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas.” [73].

“Redes neuronales artificiales son redes interconectadas masivamente en paralelo de elementos simples (usualmente adaptativos) y cono organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema biológico” [72]. Es necesario destacar que tales computadores neuronales no ejecutan las típicas instrucciones de máquina de los computadores digitales, al menos de que estén hechos para anular el comportamiento de las redes neuronales físicas. En principio, la operación de proceso básico realizada por todos los procesadores elementales es una operación análoga de transformación de sus señales de entrada [71].

En las redes neuronales biológicas, las células neuronales (neuronas) corresponden a los elementos de proceso anteriores. Las interconexiones se realizan por medio de las ramas de salida (axones) que producen un número variable de conexiones (sinapsis) con otras neuronas (o quizá con otras partes, como músculos y glándulas). Las redes neuronales son sistemas de simples elementos de proceso muy interconectados [71].

La compleja operación de las redes neuronales es el resultado de abundantes lazos de realimentación junto con no linealidades de los elementos de proceso y cambios adaptativos de sus parámetros, que pueden definir incluso fenómenos dinámicos muy complicados [71].

Una peculiaridad de las redes neuronales biológicas es su tamaño: en todo el sistema nervioso central hay del orden de 10<sup>11</sup> neuronas, pero el número de interconexiones es aún mayor, probablemente sobre las 10<sup>15</sup>. No parece posible programar las funciones de dicho sistema de acuerdo con un plan principal, teniendo en cuenta además que el tamaño y la estructura de la red están cambiando radicalmente durante y después de la niñez, cuando está ya en uso [71].

Es verdad que ciertos caracteres textuales de la red son inherentes, y durante la ontogénesis las proyecciones neuronales crecen aproximadamente hacia aquellos lugares en los cuales serán necesitados más tarde. En otras palabras, la distribución de los recursos y los caminos de comunicación más importantes están formados de acuerdo con un plan genético, mientras que el resto del programming en especial la memoria, debe ser adquirido después de nacer [71].

Programar tal red puede significar dos cosas: a) Las estructuras de interconexión entre las células son alteradas, y b) Las fuerzas de estas interconexiones son cambiadas. Parece que existen bastantes estrategias claras de cómo cambiar las fuerzas en la dirección correcta, mientras que cambios en las interconexiones son más difíciles de definir, porque suelen tener efectos radicales en el comportamiento de la red, especialmente en lo concerniente a la operación secuencial y las funciones jerárquicas [71].

Es muy difícil imaginar cómo una red tan enorme puede ser programada. Una posibilidad, con relación a los subsistemas sensoriales, podría ser que la estructura de sistema, o el proceso dinámico definido por él, de alguna forma tiende a imaginar las experiencias sensoriales de otros acontecimientos [71].

Otra función importante del sistema nervioso es definir acciones que son parte del comportamiento, y controlar el estado del organismo en relación con su entorno [71].

Mientras que las representaciones internas en las cuales está basado el comportamiento pueden ser derivadas de las entradas de forma bastante directa, las definiciones de las salidas deben estar basadas en estrategias completamente diferentes. De hecho, apenas existe otra posibilidad para programar las acciones que aplicar el principio de retroceso y empuje para alterar los mecanismos que son responsables de ellas. Algún tipo de backpropagation de la información es, por tanto, necesario [71].

Por otro lado, el significado y la calidad de las acciones deben ser juzgados no desde los movimientos inmediatos, sino desde la realización de criterios que tienen en cuenta el resultado requerido, a veces bastante indirectamente. A menudo, las acciones son solamente corregidas si se realizan con cierta frecuencia, donde el mecanismo que las provoca debe contener circuitos que definan tales frecuencias y que se cambien en relación con los resultados aprendidos. Está claro que programar las acciones es un proceso mucho más indirecto que programar las representaciones internas; el factor aleatorio no puede ser evitado [71].

Programando las funciones de entrada y de salida nos lleva sólo a una operación de comportamiento en la que el estímulo y la respuesta son consideradas muy relevantes. Ciertamente, es posible realizar autómatas bastante complejos y comportamientos necesidad-conducta de este modo. Sin embargo, se mantiene la expectativa sobre las posibilidades de que las redes neuronales puedan actuar como ordenadores para algunos problemas abstractos y también donde la computación se realice en el estado interno de la red [71].

### 7.3 Ventajas de las redes neuronales

Debido a su constitución y a sus fundamentos, las redes neuronales artificiales presentan un gran número de características semejantes a las del cerebro. Por ejemplo, son capaces de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante, etc. Esto hace que ofrezcan numerosas ventajas y que este tipo de tecnología se esté aplicando en múltiples áreas. Estas ventajas incluyen [71]:

### 7.4 Fundamentos de las redes neuronales (El modelo biológico)

La teoría y modelado de redes neuronales artificiales está inspirada en la estructura y funcionamiento de los sistemas nerviosos, donde la neurona es el elemento fundamental. Existen neuronas de diferentes formas, tamaños y longitudes. Estos atributos son importantes para determinar la función y utilidad de la neurona. La clasificación de estas células en tipos estándar ha sido realizada por muchos neuroanatomistas [71].

#### 7.4.1 Estructura de la neurona

Una neurona es una célula viva y, como tal, contiene los mismos elementos que forman parte de todas las células biológicas. Además, contienen elementos característicos que las diferencian. En

general, una neurona consta de un cuerpo celular más o menos esférico, de 5 a 10 micras de diámetro, del que salen una rama principal, el axón, y varias ramas más cortas, llamadas dendritas. A su vez, el axón puede producir ramas en torno a su punto de arranque, y con frecuencia se ramifica extensamente cerca de su extremo [71].

Una de las características que diferencia a las neuronas del resto de las células vivas, es su capacidad de comunicarse. En términos generales, las dendritas y el cuerpo celular reciben señales de entrada; el cuerpo celular las combina e integra y emite señales de salida. El axón transporta esas señales a los terminales axónicos, que se encargan de distribuir información a un nuevo conjunto de neuronas. Por lo general, una neurona recibe información de miles de otras neuronas y, a su vez, envía información a miles de neuronas más. Se calcula que en el cerebro humano existen del orden de  $10^{15}$  conexiones [71].

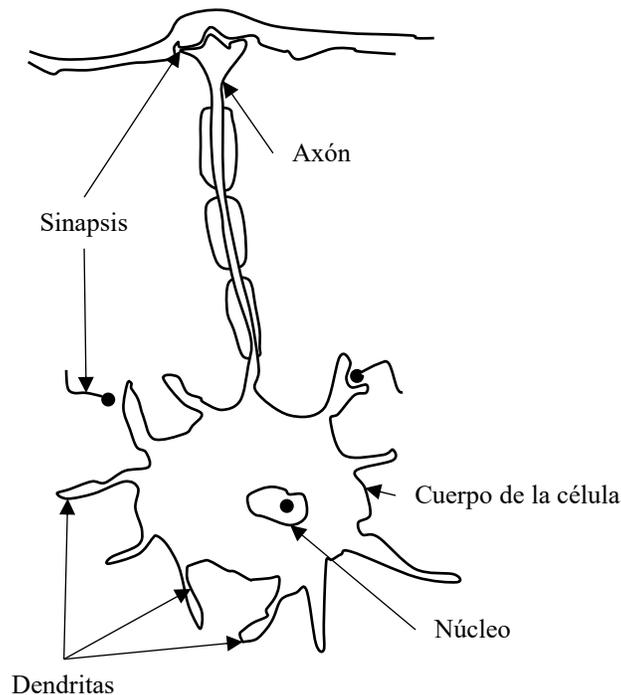


Figura 7. 1. Esquema de una red neuronal biológica

### 7.5 Naturaleza bioeléctrica de la neurona.

Las señales que se utilizan, y a las que nos estamos refiriendo, son de dos tipos distintos de naturaleza: eléctrica y química. La señal generada por la neurona y transportado a lo largo del axón es un impulso eléctrico, mientras que la señal que se transmite entre los terminales axónicos de una neurona y las dendritas de las neuronas siguientes es de origen químico; concretamente, se realiza mediante moléculas de sustancias transmisoras (*neurotransmisores*) que fluyen a través de unos contactos especiales, llamados *sinapsis*, que tienen la función del receptor y están localizados entre los terminales axónicos y las dendritas de la neurona siguiente [71].

La generación de las señales eléctricas está íntimamente relacionada con la composición de la membrana celular. Existen muchos procesos complejos relacionados con la generación de dichas

señales, sin embargo, se pueden mantener en su interior un líquido cuya composición difiere marcadamente de la composición del líquido de exterior. La diferencia más notable se da en unas 10 veces más rico en sodio que en potasio que el externo. Esta diferencia de concentración en iones sodio y potasio a cada lado de la membrana produce una diferencia de potencial de aproximadamente 70 milivoltios, negativa en el interior de la célula. Es lo que se llama *potencial de repos* de la célula nerviosa [71].

### 7.6 Red multicapa con alimentación directa (MLFF).

En las redes MLFF, las neuronas se ordenan en capas con conectividad entre las neuronas de diferentes capas. La capa que recibe las entradas se conoce como capa de entrada, y aquella que da las salidas (o vector de salida) es llamada la capa de salida. Otras capas, como estas no reciben ninguna entrada directa o contribuyen directamente a la salida, son llamadas capas ocultas. Las señales de entrada se propagan de manera gradual modificando la forma en la dirección hacia adelante, finalmente alcanzan la capa de salida. La función de activación para las neuronas en la red MLFF puede ser lineal o no lineal. La función sigmoidea es ampliamente utilizada como función de activación no lineal. Cuyas salidas caen entre 0 y 1 y se define como

$$f(x) = \frac{1}{1 + e^{-x}} \quad (7.1)$$

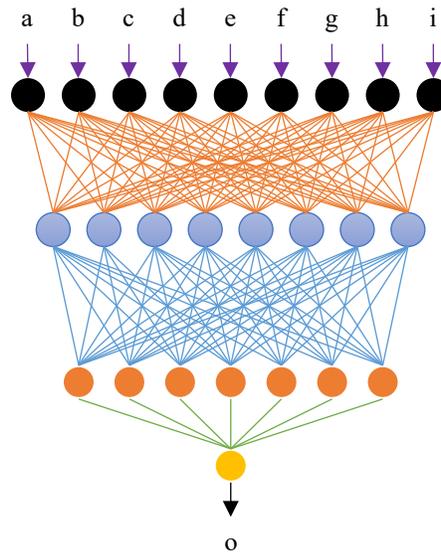
Una importante característica de esta función que la hace adecuada para usarla en conjunto con un algoritmo de aprendizaje (la modificación de los pesos se realiza en proporción al gradiente negativo de la salida) para una red MLFF que es diferenciable a lo largo de su dominio. El error para las capas ocultas se determina por propagar hacia atrás el error determinado por la capa de salida; por consiguiente la técnica se conoce como backpropagation. Durante el aprendizaje, los pesos de las neuronas se optimizan de acuerdo a la Regla Delta Generalizada (GDR), que es el algoritmo de aprendizaje para una red MLFF con retro-propagación. El error que se minimiza por el GDR es la sumatoria de los cuadrados de los errores para todas las unidades de salida, definida como:

$$E_p = \sum (y_{pk} - o_{pk})^2 \quad (7.2)$$

Para la modificación de los pesos de la capa de salida, la dirección en la que los pesos necesitan ser desplazados se determina por el gradiente negativo de  $E_p$  con respecto al peso  $w_{kj}$ . Los ajustes en el peso para cada neurona es el producto del error en la salida de la neurona, el gradiente de la salida de neurona, la entrada neta dada a la neurona y un parámetro de tasa de aprendizaje. La modificación del peso para una capa oculta se realiza en proporción al gradiente de  $E_p$  con respecto a los pesos de la capa oculta. De esta manera, cada peso actualizado en una capa oculta es dependiente de todos los términos de error de la capa de salida. Así, los errores que pueden ser determinados exactamente únicamente para la capa de salida son propagados hacia atrás a las capas ocultas. El aprendizaje de la MLFF toma lugar bajo supervisión, y un parámetro importante que tiene una influencia controlada es la tasa constante de aprendizaje. Esta decide la magnitud de

cambios en las conexiones de los pesos. Una alta tasa constante de aprendizaje tiene la ventaja de un aprendizaje más rápido, pero es posible que cause que los pesos reboten alrededor del mínimo error, impidiendo que aprendan correctamente. Por otra parte, si la tasa constante de aprendizaje es demasiado pequeña, el aprendizaje puede tomar un largo tiempo debido al lento descenso a lo largo de la superficie del error, lo que puede ser favorable para que la red pueda encontrar el mejor mínimo error y, por lo tanto, un aprendizaje más exacto [71].

Para estimar la radiación media global diaria a partir de datos mínimos disponibles sobre la posición para cualquier ubicación, una red multicapa con pre-alimentación se entrenó y se probó para su capacidad para generalizar e interpolar. La estructura de RNA seleccionada **Figura 7.2** es una pre-alimentada, red jerárquica totalmente conectada consistente de una capa de entrada, dos capas ocultas y una capa de salida. La primera capa oculta tiene ocho neuronas, y la segunda capa oculta tiene siete neuronas. Hay una única neurona de salida. La Retro-propagación interactiva con el algoritmo GDR ha sido implementada para determinar los errores para las neuronas de la capa oculta y subsecuente modificación de peso de acuerdo al GDR [71].



**Figura 7.2.** RNA, estructura utilizada para obtener la estimación de la radiación solar horaria

### 7.7 Red no lineal autorregresiva con entradas exógenas (NARX)

Considerando una red recurrente con una sola entrada y una sola salida, cuyo comportamiento esta descrito por las ecuaciones de estado (7.1) y (7.2). Dado este modelo espacio estado, se desea modificarlo dentro de un modelo de entrada – salida como una representación equivalente de la red recurrente [74].

$$x(n + 1) = \varphi(W_a x(n) + w_b u(n)) \quad (7.1)$$

$$y(n) = c^T x(n) \quad (7.2)$$

Utilizando las [ecuaciones 7.1 y 7.2](#), se puede demostrar fácilmente que la salida  $y(n + q)$  es expresable en términos del estado  $x(n)$  y el vector de entrada  $u_q(n)$  como sigue:

$$y(n + q) = \Phi(x(n), u_q(n)) \quad (7.3)$$

Donde  $q$  es la forma dimensional del espacio estado, y  $\Phi: \mathbb{R}^{2q} \rightarrow \mathbb{R}$ . Siembre que la red recurrente sea observable, podemos utilizar el teorema de observabilidad local para escribir

$$x(n) = \Psi(y_p(n), u_{q-1}(n)) \quad (7.4)$$

Donde  $\Psi: \mathbb{R}^{2q-1} \rightarrow \mathbb{R}^q$ . Por lo tanto, sustituyendo la ecuación (7.4) en (7.3), obtenemos

$$\begin{aligned} y(n + q) &= \Phi(\Psi(y_p(n), u_{q-1}(n)), u_q(n)) \\ &= F(y_q(n), u_q(n)) \end{aligned} \quad (7.5)$$

Donde  $u_{q-1}(n)$  está contenida  $u_q(n)$  como sus primeros elementos ( $q - 1$ ), y el mapeo no lineal  $F: \mathbb{R}^{2q} \rightarrow \mathbb{R}$  toma en consideración  $\Phi$  y  $\Psi$ . Utilizando las definiciones de  $y_p(n)$  y  $u_q(n)$  dadas en las [ecuaciones 7.6 y 7.7](#):

$$u_{q-1}(n) = [u(n), u(n + 1), \dots, u(n + q - 2)]^T \quad (7.6)$$

$$y_p(n) = [y(n), y(n + 1), \dots, y(n + q - 1)]^T \quad (7.7)$$

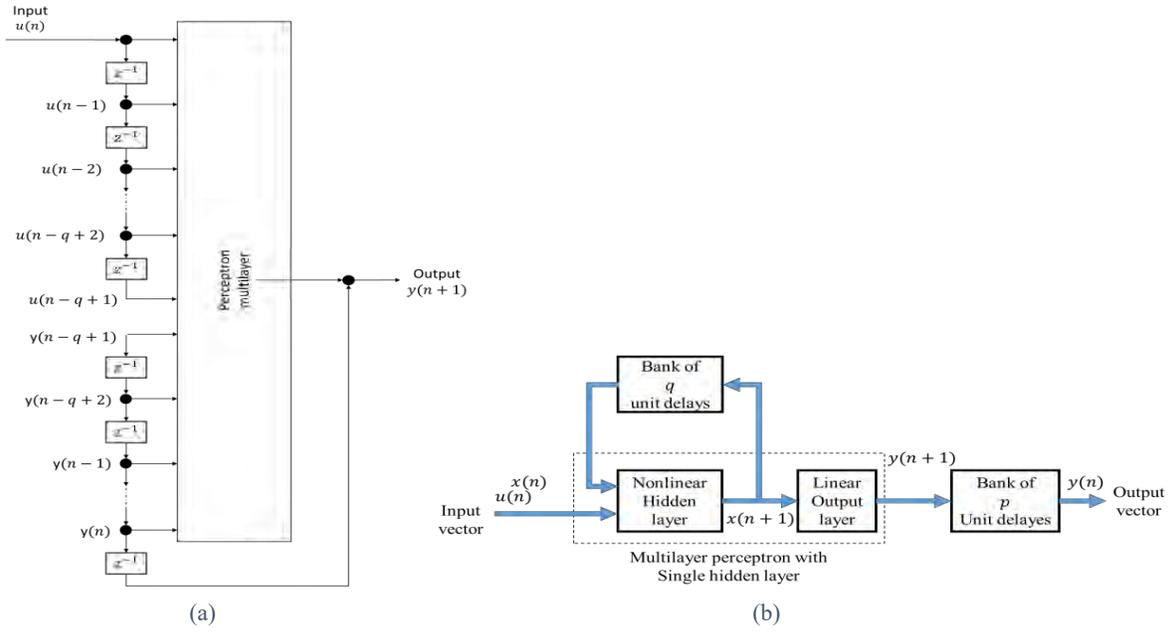
podemos escribir la [ecuación 7.5](#) en la forma expandida:

$$y(n + q) = F(y(n + q - 1), \dots, y(n), u(n + q - 1), \dots, u(n))$$

Remplazando  $n$  con  $n - q + 1$ , podemos escribir de manera equivalente:

$$y(n + 1) = F(y(n), \dots, y(n - q + 1), u(n), \dots, u(n - q + 1)) \quad (7.8)$$

Indicado en palabras, algunos de los mapeos no lineales  $F: \mathbb{R}^{2q} \rightarrow \mathbb{R}$  existen con lo cual el valor presente de la salida  $y(n + 1)$  se define de forma única en términos de sus valores pasados  $y(n), \dots, y(n - q + 1)$  y los valores presentes y pasados de la entrada  $u(n), \dots, u(n - q + 1)$ . Para que esta representación de entrada-salida sea equivalente al modelo espacio-estado de las [ecuaciones 7.1 y 7.2](#), la red recurrente debe de ser observable. La representación práctica de esta equivalencia es que el modelo NARX de la Fig. 3, con su retroalimentación global limitada hacia la neurona de salida, esta de hecho disponible para simular plenamente el correspondiente modelo espacio – estado recurrente de la Fig. 2 (asumiendo que  $m = 1$  y  $p = 1$ ) sin diferencias entre su comportamiento de entrada – salida [74].



**Figura 7.3.** (a) Red neuronal no lineal autorregresiva con entradas externas, (b) Perceptrón Múltiplo con una capa oculta [74]

# Capítulo

## 8

### Modelado de sistemas fotovoltaicos

En el siglo actual, el mundo está experimentando una creciente necesidad de fuentes adicionales de energía, así como de reducir la dependencia sobre las fuentes convencionales, y la energía fotovoltaica (FV) puede ser la respuesta a esta necesidad. Las celdas FV están siendo utilizadas en aplicaciones espaciales y terrestres donde son económicamente competitivas como fuentes alternas. Además, la industria FV ha demostrado tasas altas de crecimiento en años recientes [75]. La importancia de incluir este tipo de sistemas a la red ha llevado a la elaboración de diversos estudios y trabajos de investigación. En [76] se lleva a cabo una revisión de las técnicas utilizadas para analizar el desempeño en los módulos FV y la dependencia del desempeño que estos tienen sobre la temperatura. [77] desarrollan estudios de simulación de un sistema de potencia, con el objetivo de introducir los principales componentes, modos de operación/protección, y control de capas/esquemas de mediana y alta potencia en sistemas FV.

[78] elaboran un análisis y modelado del desempeño del modelo de celdas solares FV utilizando los módulos incorporados en ambiente MATLAB/SIMULINK. El principal objetivo fue encontrar curvas características no lineales de intensidad versus tensión ( $I - V$ ) y potencia versus tensión ( $P - V$ ) para módulos FV Photonix 150W y 140W. [79] presentan un método sencillo y preciso para el modelado de arreglos FVs. El método se utiliza para obtener los parámetros del modelo de arreglo utilizando información de la hoja de fabricante.

En este trabajo de investigación se desarrolla un software para obtener las curvas características  $V - I$  para módulos FV ISF-240 y para obtener la potencia de salida producida por el generador FV utilizando series de tiempo de irradiación solar y temperatura ambiente.

#### 8.1 Modelado de Sistemas FV.

##### 8.1.1 Modelo de la celda fotovoltaica (FV).

La **Figura 8.1** muestra el circuito equivalente de una celda solar, el cual consiste en una fuente de corriente en paralelo con un diodo. La salida de la fuente de corriente es directamente proporcional a la luz que incide sobre la celda. En la oscuridad, la celda solar no es un dispositivo activo, ya que funciona como un diodo, es decir, una unión p-n. No se produce corriente ni tensión. Sin embargo, si la celda está conectada a una fuente externa se genera una corriente llamada corriente de oscuridad. Por lo tanto, el diodo determina las características  $V - I$  de la celda.

La corriente total del circuito de la **Figura 8.1** se define como la diferencia entre la corriente debida a la radiación solar  $I_l$  y la corriente de oscuridad  $I_D$ . Esta relación es definida como [80],

$$I_c = I_l - I_D \quad (8.1)$$

$$I_D = I_o \left( e^{\frac{V_c}{V_T}} - 1 \right) \quad (8.2)$$

Con,

$$V_T = \frac{q}{kT_{ck}}$$

En donde,

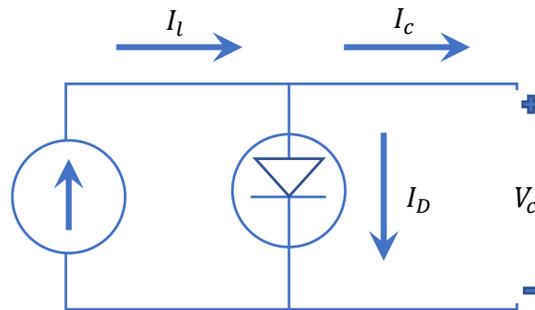
$I_D$	Corriente de oscuridad, A
$I_c$	Corriente de la celda, A
$I_o$	Corriente de saturación del diodo, A
$V_c$	Voltaje de la celda, V
$V_T$	Potencial térmico, V
$q$	$1.6 \times 10^{-19}$ , coulomb
$k$	$1.38 \times 10^{-23}$ , J/K
$T_{ck}$	Temperatura de la celda, K

La **Figura 8. 2** muestra el circuito equivalente utilizado para el modelado del generador fotovoltaico (GFV). Esta configuración añade una resistencia en serie  $R_s$  que representa las pérdidas internas debidas al flujo de corriente.

La corriente de la celda correspondiente a la **Figura 8. 2** se define como [80]

$$I_c = I_{sc} - I_o \times \left( e^{\frac{V_c + I_c R_s}{nV_T}} - 1 \right) \tag{8.3}$$

$I_{sc}$	Corriente de corto circuito de la celda, A
$R_s$	Resistencia serie, $\Omega$
$n$	Constante de no idealidad del diodo.

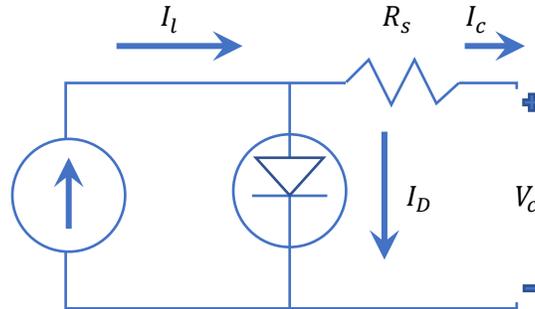


**Figura 8. 1.** Modelo simplificado de la celda solar

Con el propósito de modelar un módulo FV se considera que las características de corriente y voltaje de cada celda del generador son uniformes. Esta consideración es importante por dos razones principales.

- Asegura que los diodos de derivación se polarizan inversamente y son tratados como circuitos abiertos. Además, el diodo de bloqueo está polarizado directamente y se trata como un corto circuito.

- Garantiza que la expresión exponencial (8.3) puede escalarse de forma lineal basada en las conexiones serie y paralelo de las celdas debido a que los parámetros en el exponente ( $V_c$ ,  $I_c$ ,  $T_{cK}$ ) serán idénticos para cada celda [80].



**Figura 8. 2.** Modelo de la celda solar con resistencia serie

Teniendo en cuenta que en una conexión en serie los voltajes se suman y en una conexión en paralelo las corrientes se suman, entonces se aplican las siguientes relaciones,

$$\begin{aligned} I_M &= N_p I_c \\ I_{scM} &= N_p I_{sc} \\ V_M &= V_c N_s \\ V_{ocM} &= V_{oc} N_s \end{aligned} \quad (8.4)$$

En donde,

$N_s$	Número de celdas conectadas en serie
$N_p$	Número de celdas conectadas en paralelo
$I_{scM}$	Corriente de corto circuito del módulo, A
$I_M$	Corriente del módulo, A
$V_M$	Voltaje del módulo, V
$V_{ocM}$	Voltaje de circuito abierto del módulo, V
$V_{oc}$	Voltaje de circuito abierto de la celda, V

Sustituyendo (8.4) en (8.3) se tiene,

$$I_M = I_{scM} - N_p I_o \left( e^{\frac{V_M + I_M R_{sM}}{n N_s V_T}} - 1 \right) \quad (8.5)$$

## 8.2 Modelado del generador fotovoltaico

El generador fotovoltaico (GFV) está constituido por un conjunto de módulos fotovoltaicos conectados en serie y en paralelo. El conjunto de módulos fotovoltaicos genera la potencia

requerida. A continuación, se enumeran los pasos básicos para determinar los parámetros del módulo.

Una vez que se determina el esquema de conexión, la ecuación característica del módulo se escala para describir el comportamiento de todo el generador. Esto se hace de acuerdo con (8.4) y (8.5), de esta manera se tiene,

$$\begin{aligned} I_{scG} &= I_{scM} N_{pM} \\ V_{ocG} &= V_{ocM} N_{sM} \\ P_{nG} &= N_{sM} N_{pM} P_{nM} \end{aligned} \quad (8.6)$$

En donde  $I_{scG}$ ,  $V_{ocG}$  y  $P_{nG}$  representan la corriente de corto circuito, el voltaje de circuito abierto y la potencia del generador, respectivamente.

El número de celdas es el producto de celdas en serie del módulo  $N_s$  y el número de módulos en cadena  $N_{sM}$ . El número de conexiones en paralelo de las celdas es simplemente el número de cadenas paralelas de módulos:

$$\begin{aligned} N_{sG} &= N_s N_{sM} \\ N_{pG} &= N_{pM} \end{aligned} \quad (8.7)$$

Dadas las condiciones anteriores, la ecuación característica V-I que se utiliza para modelar el GFV se define con,

$$I_G = I_{scG} - N_{pG} I_o \left( e^{\frac{V_G + I_G R_{sG}}{n N_{sG} V_T}} - 1 \right) \quad (8.8)$$

El parámetro  $I_o$  se determina resolviendo la [Ecuación 8.8](#) para las siguientes condiciones de circuito abierto,

$$\begin{aligned} I_G &= 0 \\ V_G &= V_{ocGE} \end{aligned}$$

Despejando  $I_o$  de (8.8),

$$I_o = \frac{I_{scGE}}{N_{pG} \left( e^{\frac{V_{ocGE}}{n N_{sG} V_T}} - 1 \right)} \quad (8.9)$$

En donde  $I_{scGE}$  y  $V_{ocGE}$  son la corriente de corto circuito y tensión de circuito abierto a condiciones ambientales para el generador. La resistencia  $R_{sG}$  del generador se define como [80],

$$R_{sG} = \frac{V_{ocG}}{I_{scG}} - \frac{P_{sG}}{FF_0 I_{scG}^2} \quad (8.10)$$

Con,

$$FF_0 = \frac{v_{oc} - \ln(v_{oc} - 0.72)}{1 + v_{oc}} \quad (8.11)$$

$$v_{oc} = \frac{V_{ocG}}{N_{SG} n V_T}$$

En donde,

$FF_0$  Factor de relleno del generador sin resistencia serie  
 $v_{oc}$  Valor normalizado para el voltaje de circuito abierto

### 8.2.1 Modelo dependiente de la temperatura

La corriente de cortocircuito del generador fotovoltaico con dependencia de la temperatura se define con [80],

$$I_{scGE} = \frac{I_{scGn}}{1000} + \left( \frac{\partial I_{scG}}{\partial T_c} \right) (T_c - T_{c0}) \quad (8.12)$$

En donde,

$T_c$  Temperatura de la celda, °C  
 $T_{c0}$  Temperatura de la celda a condiciones de operación nominal (usualmente 25 °C)  
 $I_{scGE}$  Corriente de corto circuito del generador para condiciones ambientales dadas  
 $\left( \frac{\partial I_{scG}}{\partial T_c} \right)$  Coeficiente de temperatura para la corriente de corto circuito  
 $I_{scGn}$  Corriente de corto circuito nominal del generador (máximo punto de potencia a condiciones de operación nominal)

El voltaje de circuito abierto del generador dependiente de la temperatura se define con,

$$V_{ocGE} \approx V_{ocGn} + \left( \frac{\partial V_{ocG}}{\partial T_c} \right) (T_c - T_{c0}) + \frac{kT_{cK}}{q} \ln \left( \frac{I_{scGE}}{I_{scGn}} \right) \quad (8.13)$$

$\left( \frac{\partial V_{ocG}}{\partial T_c} \right)$  Coeficiente de temperatura del voltaje de circuito abierto  
 $V_{ocGE}$  Voltaje de circuito abierto para condiciones ambientales dadas

A medida que la temperatura de las celdas aumenta, la capacidad de voltaje del material semiconductor en la celda se ve disminuida y, por lo tanto, el voltaje a través de la celda disminuye. La temperatura de la celda se puede calcular con las medidas de temperatura ambiente y la temperatura NOCT (Nominal Operating Cell Temperature), el cual es un dato que proporciona el

fabricante en la hoja de datos del módulo. De esta manera, se tiene que la temperatura de la celda y del módulo estará dada por,

$$T_c = T_a + \frac{NOCT - 20}{800} G \tag{8.14}$$

En donde,

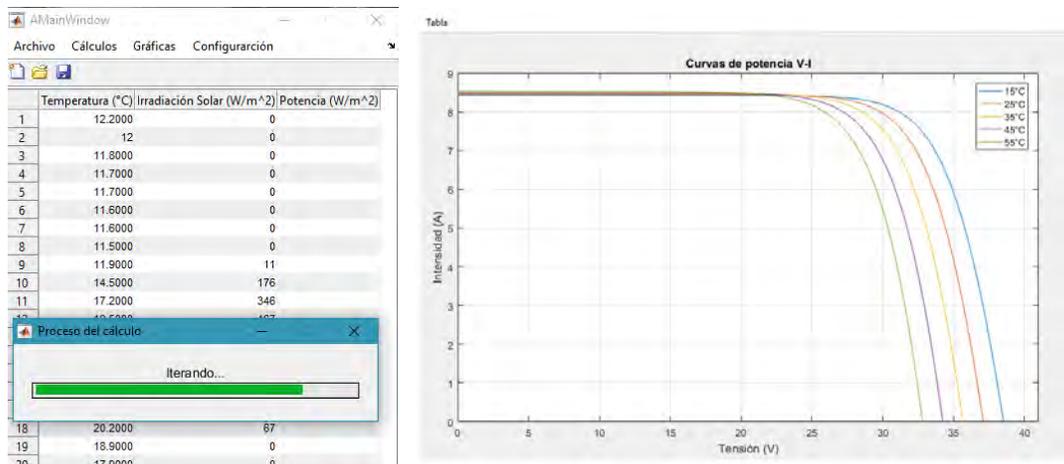
- NOCT* Temperatura nominal de operación proporcionada por el fabricante
- $T_c$  Temperatura de la celda, °C
- $T_a$  Temperatura ambiente
- $G$  Irradiancia,  $W/m^2$

### 8.3 Resultado de la potencia generada utilizando el software.

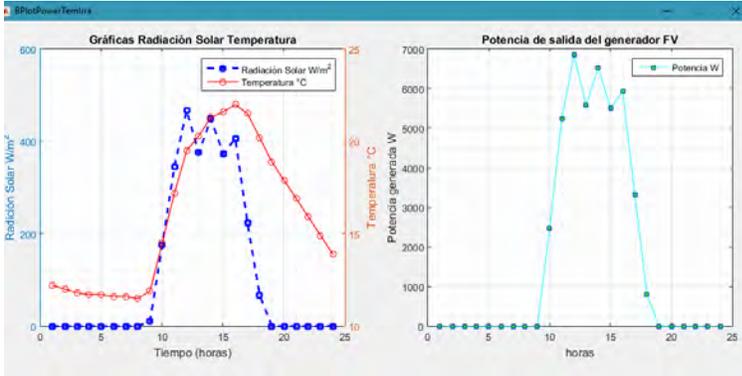
En este apartado se presentan los resultados de la potencia generada de acuerdo a los datos de temperatura y radiación solar que se ingresan en el programa. Los resultados corresponden a un GFV, con 5 módulos conectados en paralelo, 14 en conectados en serie, con paneles ISF-240 con una  $I_{sc} = 8.45A$  y un  $V_{oc} = 37.1V$  con una potencia máxima de  $P_{Mpp} = 240W$ .

En la **Figura 8. 3a** se muestra la ventana principal la cual consta de un menú principal y una barra de herramientas para facilitar la interacción con el usuario, en la primera columna se carga la serie de tiempo de temperaturas, en la segunda columna se muestran las series de tiempo de radiación solar, mientras que la tercera columna muestra la potencia calculada en el GFV debido a la temperatura y radiación de entrada.

La **Figura 8. 3b** presenta las curvas características V- I para cinco temperaturas 15, 25, 35, 45 y 55 °C, gráficas calculadas por el programa y que coinciden con las proporcionadas en la hoja de datos. En la **Figura 8. 3a** se muestran las gráficas de temperatura y radiación solar, junto con la gráfica de la potencia que genera el módulo, mientras que la **Figura 8. 3b** muestra las opciones que el usuario puede modificar de acuerdo a las características del módulo que se necesite analizar, esta ventana cuenta con tres partes principales: Módulo, Conexión y Curva I – V.



**Figura 8. 3.** (a) Ventana principal y (b) gráficas de curvas de potencia.



(a)

The image shows a software dialog box titled 'COpciones' with a tab labeled 'Módulo' and sub-tabs 'Conexión' and 'Curva I-V'. Under 'Propiedades del Módulo', there are several input fields with values and units:

IscM:	8.45	A
VocM:	37.1	V
dIdT:	0.0294	K/A
dVdT:	-0.387	KV
Pmpp:	240	W
n:	1.6	
NOCT:	47	°C

An 'Aceptar' button is located at the bottom right of the dialog.

(b)

**Figura 8.4.** Gráfica temperatura y Radiación solar (b) panel de opciones

# Capítulo

## 9

### Desarrollo de modelos de pronóstico

En este capítulo se muestran los resultados de los modelos de pronóstico que se utilizaron para prever la radiación solar. Los datos meteorológicos de series de tiempo se obtuvieron de la estación de Temixco, Morelos, la cual proporciona datos de radiación solar ( $R_S$ ) en  $W/m^2$ , temperatura ( $T$ ) en  $^{\circ}C$ , humedad relativa ( $H_R$ ) en %, presión atmosférica ( $P$ ) en Pa y velocidad del viento ( $V_V$ ) en  $m/s^2$ .

#### 9.1 Análisis estadístico de la serie de tiempo de radiación solar

Como se observa en la **Tabla 9. 1**, los meses más variables son julio y septiembre, y los meses con menos variación son marzo y abril. Mientras que los meses que registran el valor más alto de  $R_S$  son mayo y agosto.

**Tabla 9. 1.** Medidas estadísticas para la serie de tiempo de Temixco, 2010

2010	Máximo $W/m^2$	Mínimo $W/m^2$	Media $W/m^2$	Des. Est. $W/m^2$	Sesgo	Curtosis	CV %
A anual	1059	1	430.9	312.5	0.09	1.63	73%
enero	927	2	355.0	264.9	0.25	1.75	75%
febrero	960	1	419.0	303.3	0.10	1.54	72%
marzo	993	1	510.1	329.5	-0.15	1.57	65%
abril	1022	1	522.0	333.0	-0.16	1.60	64%
mayo	1059	1	467.2	337.9	0.01	1.49	72%
junio	959	1	454.1	333.1	0.03	1.49	73%
julio	1006	1	366.2	307.7	0.43	1.83	84%
agosto	1059	1	425.4	312.5	0.16	1.71	73%
septiembre	1011	1	397.8	307.7	0.23	1.68	77%
octubre	911	3	452.9	310.1	-0.15	1.52	68%
noviembre	847	1	397.2	284.1	-0.08	1.51	72%
diciembre	761	1	394.6	257.0	-0.21	1.57	65%

En la **Tabla 9. 2** se presentan las principales medidas estadísticas de la radiación solar correspondientes al año 2011, como podemos ver los meses más variables son julio y septiembre, mientras que los menos variables son enero y febrero. Estos valores son un indicativo de que es muy probable que el modelo tienda a ser menos precisos en los meses con un CV muy alto, mientras que deberá ser más precisos en aquellos con CV más bajo.

**Tabla 9. 2.** Medidas estadísticas para la serie de tiempo de Temixco, 2011

2011	Máximo W/m <sup>2</sup>	Mínimo W/m <sup>2</sup>	Media W/m <sup>2</sup>	Des. Est. W/m <sup>2</sup>	Sesgo	Curtosis	CV %
Anual	1056	1	453.59	316.08	-0.02	1.59	69.68%
enero	821	5	427.42	266.33	-0.25	1.63	62.31%
febrero	884	11	477.96	291.11	-0.26	1.62	60.91%
marzo	965	1	503.35	320.59	-0.21	1.57	63.69%
abril	966	1	489.77	337.44	-0.12	1.48	68.90%
mayo	1056	1	465.47	326.98	-0.07	1.48	70.25%
junio	1021	1	450.95	331.22	0.10	1.58	73.45%
julio	1026	1	423.64	330.31	0.25	1.66	77.97%
agosto	1056	1	476.44	339.96	0.01	1.59	71.36%
septiembre	1006	1	445.89	330.09	0.11	1.57	74.03%
octubre	942	1	453.82	319.07	-0.10	1.51	70.31%
noviembre	868	1	410.96	293.41	-0.08	1.50	71.40%
diciembre	800	1	415.60	268.01	-0.23	1.56	64.49%

## 9.2. Función de autocorrelación y autocorrelación parcial

Como se explicó en la sección 6.2, la función de autocorrelación y autocorrelación parcial pueden ser de gran ayuda para la identificación de modelos potenciales de predicción. En la **Figura 9. 1**, se muestra (a) la función de autocorrelación serie de tiempo original, (b) función de autocorrelación parcial serie de tiempo original, (c) función de autocorrelación serie de tiempo con diferencia de 24 retrasos, (d) función de autocorrelación parcial con diferencia de 24 retrasos.

A partir de la **Figura 9. 1(a)** y **(b)**, podemos llegar a la conclusión de que el modelo predominante será uno del tipo ARIMA estacional con una temporalidad de 24 rezagos, en la parte estacional y en la parte estacional los modelos posibles serán deferentes combinaciones de ARIMA(p,d,q), por lo que los modelos a ensayar quedan de la siguiente manera ARIMA(p,d,q)(Q,D,Q)<sub>24</sub>.

El primer modelo que se utilizó fue un modelo ARIMA estacional, este tipo de modelos son univariados, lo que significa que realizan pronósticos basándose en valores pasados de la serie de tiempo. Fue desarrollado utilizando una serie de tiempo de radiación solar del año 2010 medida en Temixco, Morelos, como se menciona anteriormente y se pronosticó el año 2011. El primer paso es obtener las medidas principales estadísticas, es decir: Máximo de radiación solar, mínimo, media o promedio, desviación estándar, sesgo, curtosis y coeficiente de variación. (ver **Tabla 9. 1**)

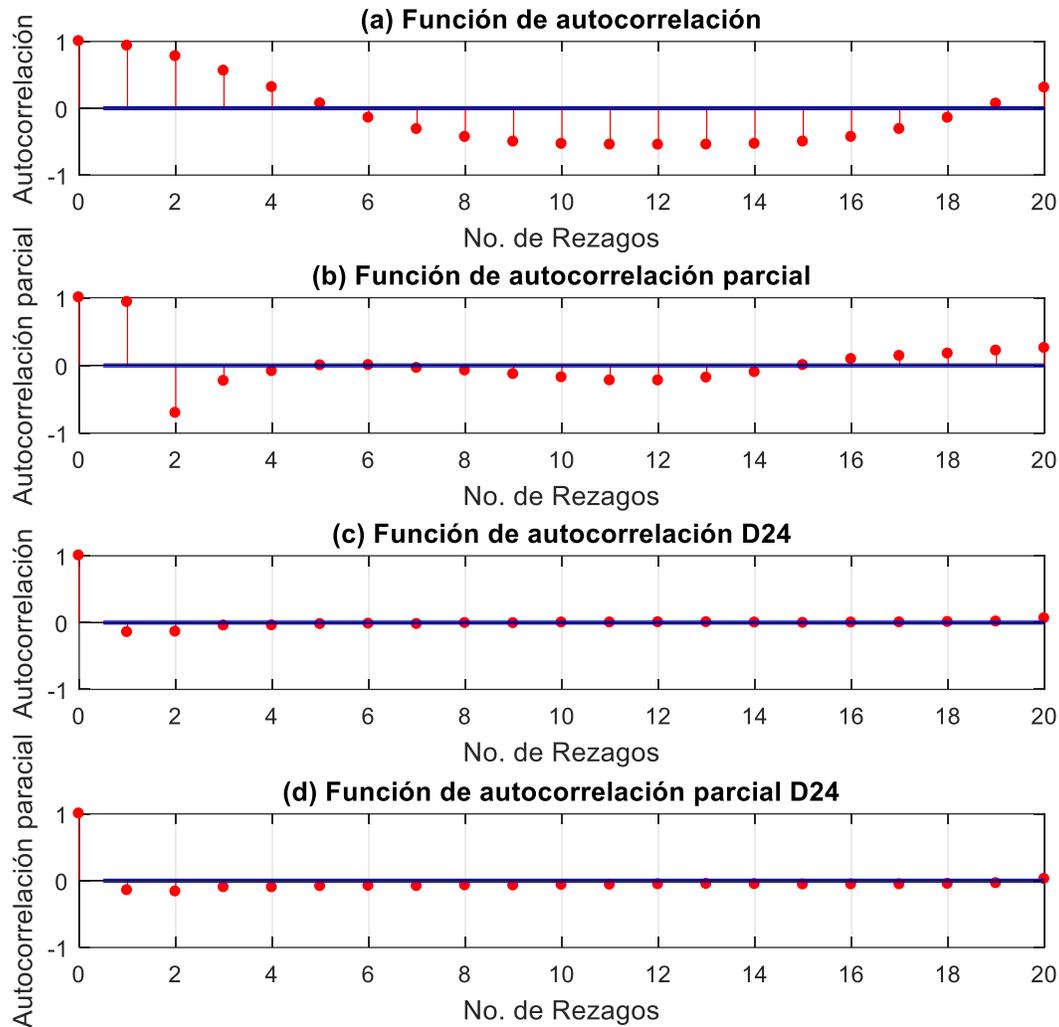


Figura 9. 1 (a) Función de autocorrelación serie de tiempo original, (b) Función de autocorrelación parcial serie de tiempo original, (c) Función de autocorrelación diferenciada 24 rezagos

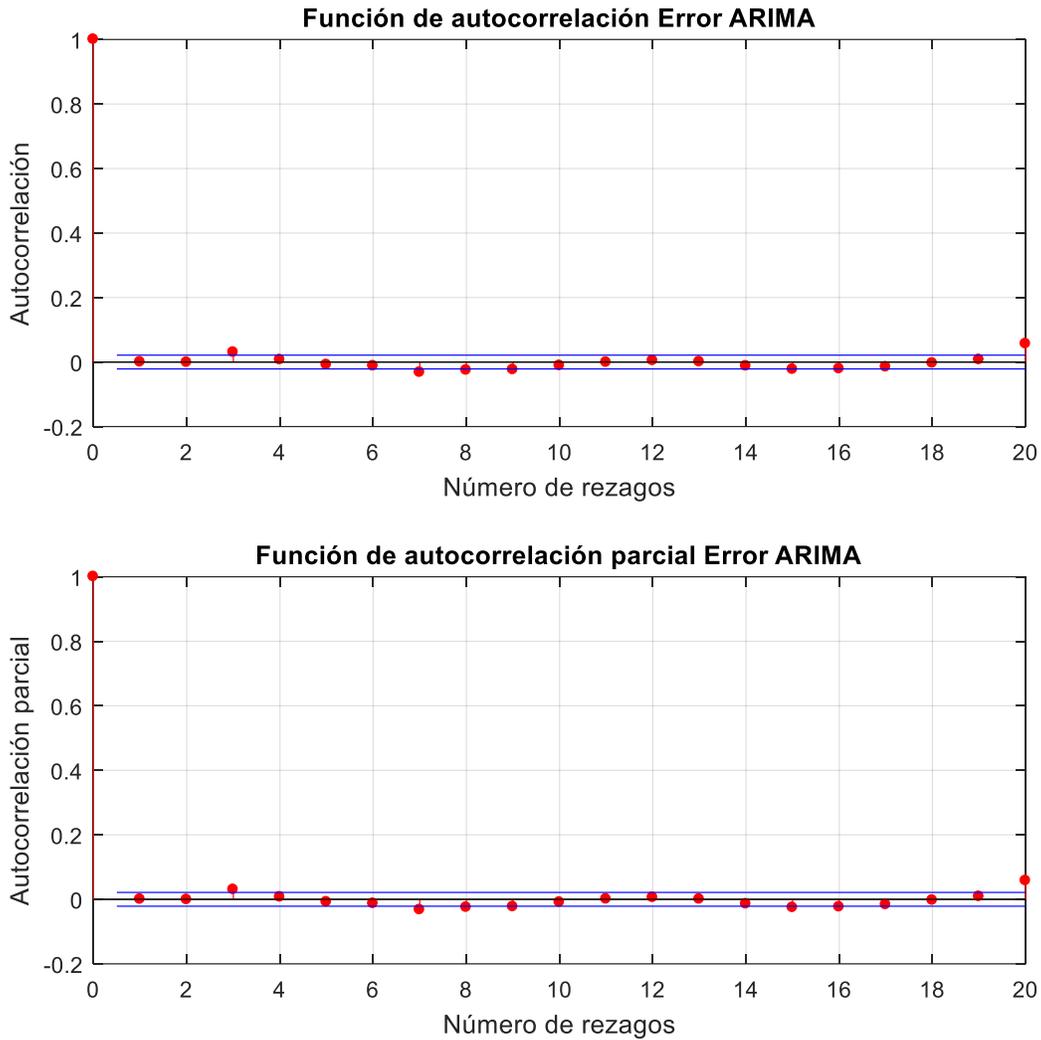
### 9.3 Modelo estacional ARIMA

Una vez que se calcularon los datos estadísticos más relevantes se procede a desarrollar el modelo de predicción, para esto se probaron 150 modelos, resultando que el mejor modelo fue un  $ARIMA(1,0,1)(24,0,24)_s$ .

#### 9.3.1 Análisis y monitoreo de los errores modelo ARIMA

Se obtiene el pronóstico de la serie de tiempo, se calcula el error, al cual le aplicamos las pruebas de función de autocorrelación y función de autocorrelación parcial. Como se explica en el capítulo 6, un buen modelo de pronóstico es aquel cuyo análisis de los errores de predicción son ruido blanco. Esto se puede determinar a partir de la función de autocorrelación y autocorrelación parcial

analizando los errores generados al sustraer los datos generados por el modelo a los datos reales (error de pronóstico).



**Figura 9. 2.** (a) Función de autocorrelación errores de la curva de ajuste (b) función de autocorrelación parcial para los mismos errores

A partir de la **Figura 9. 2**, podemos concluir que el modelo es adecuado para utilizarlo para pronósticos.

En la **Tabla 9. 3** se muestra las principales medidas estadísticas de desempeño para los datos de ajuste, como podemos ver el  $ME \approx 0$ , este es otro indicativo de que se obtuvo un buen modelo.

**Tabla 9. 3.** Resultados de las pruebas de desempeño, Curva de ajuste

Precisión	ME	MSE	RMSE	MAE
516.6	0.038	26.49	3.21E+03	56.65

### 9.3.2 Uso del modelo ARIMA para la predicción 24 horas adelante

Con el mejor modelo se pronóstica el año siguiente, cada 24 horas, ya que se cuenta con las series de tiempo de los dos años, la validación del modelo se realiza generando el modelo con la serie del año 2010 como se explicó anteriormente, se simula que no se cuenta con los datos del 2011, y que se pronosticará el 1 de enero del 2011, utilizando los últimos datos de diciembre de 2010.

Teniendo los datos reales del 1 de enero de 2011, estos junto con algunos datos de diciembre de 2010 se utilizan para predecir los datos de  $R_S$  del 2 de enero del 2011, estos pasos se repiten hasta predecir los 365 días del año 2011.

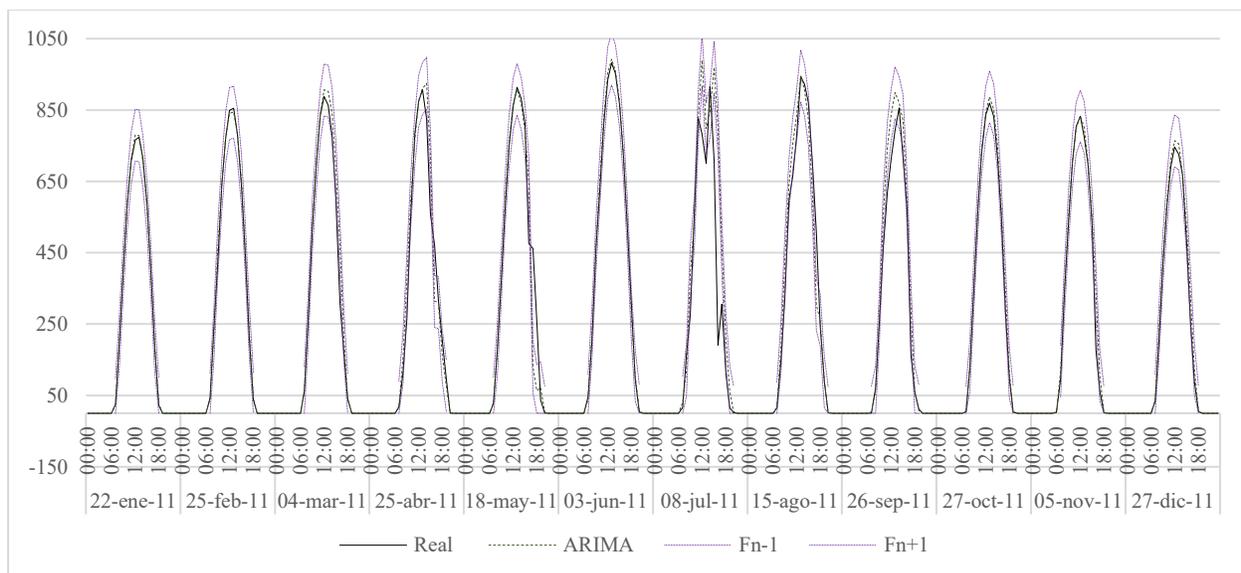
Se eligió de forma aleatoria un día de cada mes del año pronosticado, los resultados se presentan en la **Figura 9.3**.

Los días resultantes fueron: el 22 de enero, 25 de febrero, 4 de marzo, 25 de abril, 18 de mayo, 3 de junio, 8 de julio, 15 de agosto, 26 de septiembre, 27 de octubre, 5 de noviembre y 27 de diciembre.

**Tabla 9.4.** Resultado ARIMA de las pruebas de desempeño, Curva de ajuste, Datos predichos, Un día de cada mes

	Precisión	ME	MSE	RMSE	MAE
<b>Curva de ajuste 2010</b>	516.6	0.038	26.49	3.21E+03	56.65
<b>Pronóstico 2011</b>	776.3	-0.001	34.39	7.77E+03	88.14
<b>Un día para cada mes</b>	608.8	11.954	21.93	3.66E+03	60.52

En la **Tabla 9.4**, se muestra el resultado de las pruebas de desempeño para la curva de ajuste, los datos predichos y para un día de cada mes del año pronosticado, elegidos de forma aleatoria y los cuales corresponden a las fechas que se muestran en la gráfica de la **Figura 9.3**, y mencionadas previamente.



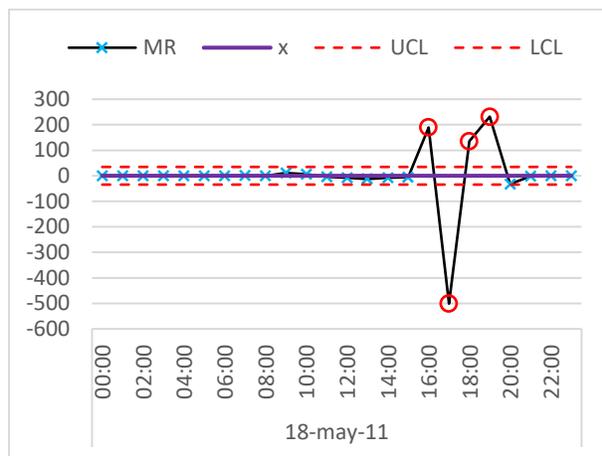
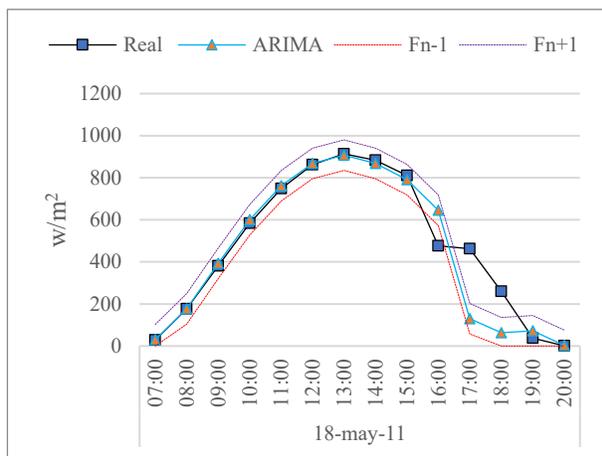
**Figura 9.3.** Pronóstico modelo ARIMA de 1 día de cada mes del año 2011, elegido de forma aleatoria

Con la finalidad de realizar un análisis más profundo, también tabulan las pruebas de desempeño para cada día del mes de los datos que se muestran en la gráfica de la **Figura 9.3**, los resultados se pueden consultar en la **Tabla 9.5**.

**Tabla 9.5.** Pruebas de desempeño individuales para un día de cada mes

Mes	Precisión	ME	MAE	MSE	RMSE
22-ene-11	14	1.1	2.8	2.33E+01	5
25-feb-11	35	1.9	5.7	1.09E+02	10
04-mar-11	156	17.2	17.7	1.66E+03	41
25-abr-11	163	8.9	29.2	3.29E+03	57
18-may-11	332	-13.0	34.4	7.50E+03	87
03-jun-11	32	4.4	7.7	1.39E+02	12
08-jul-11	609	61.4	73.0	2.30E+04	151
15-ago-11	202	-0.2	29.5	3.20E+03	57
26-sep-11	125	29.5	30.6	3.19E+03	56
27-oct-11	57	8.6	8.6	2.52E+02	16
05-nov-11	157	12.4	12.7	1.18E+03	34
27-dic-11	74	11.3	11.3	4.65E+02	22

De acuerdo a los resultados de la **Tabla 9.5** los peores pronósticos se obtuvieron para los días 15 de agosto y 5 de noviembre de 2011 y los mejores, para los días 25 de febrero y 27 de octubre de 2011.



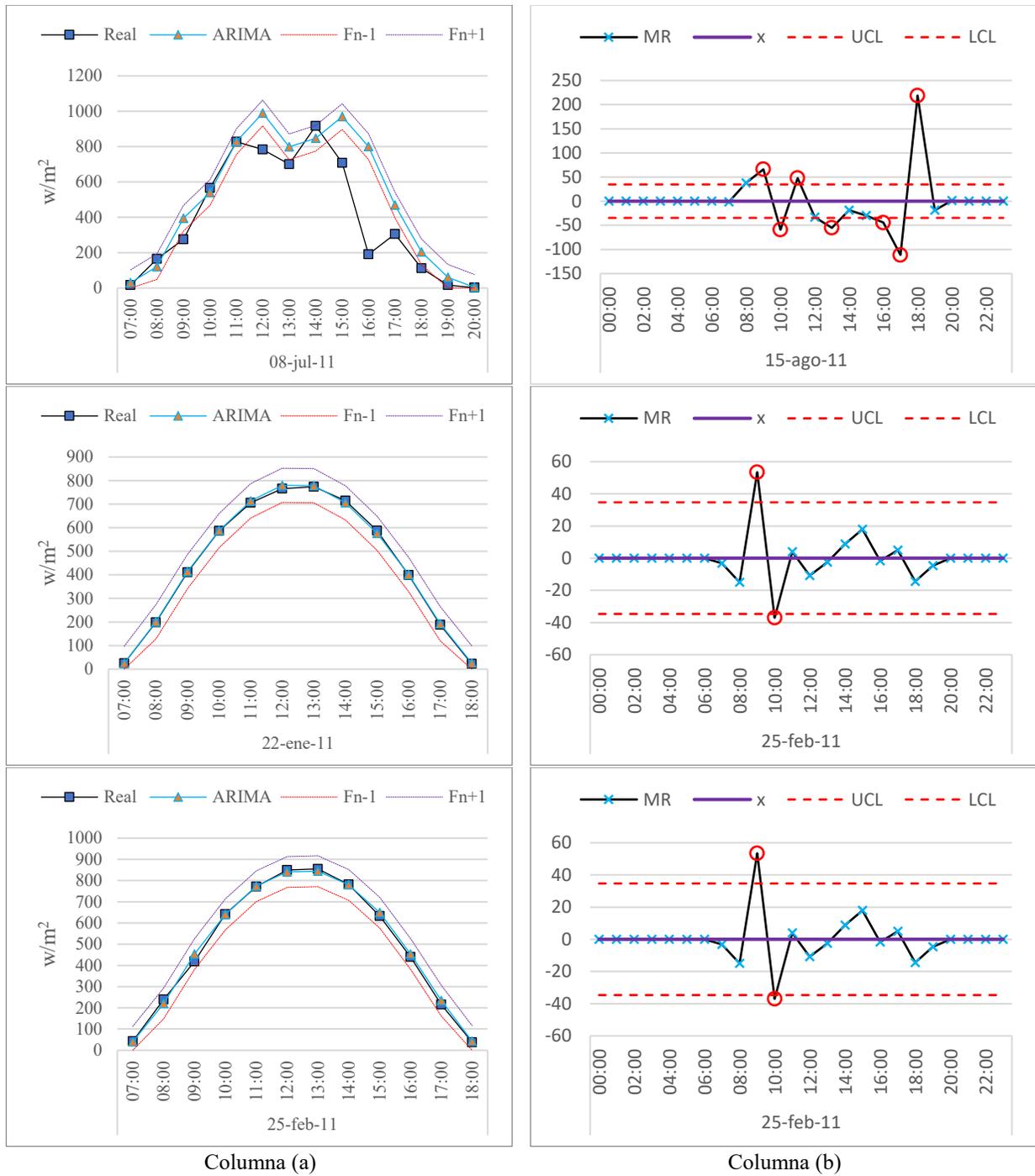


Figura 9. 4. Columna (a) Resultados del pronóstico, Columna (b) caja de control. Modelo ARIMA

La **Tabla 9. 6** muestra los resultados de las pruebas de desempeño aplicadas a cada uno de los meses. De acuerdo a la **Tabla 9. 6** los peores resultados se obtuvieron en los meses de junio y julio del 2011, mientras que los mejores resultados son aquellos para enero y diciembre del mismo año.

**Tabla 9. 6.** Resultados ARIMA de las pruebas de desempeño para cada mes de pronóstico

Mes	Precisión	ME	MAE	MSE	RMSE
Ene	263	-1.3	15.2	1.46E+03	38
Feb	353	-0.5	14.2	1.68E+03	41
Mar	321	-1.1	14.9	1.49E+03	39
Abr	653	0.7	25.1	4.41E+03	66
May	461	0.8	29.4	4.53E+03	67
Jun	743	3.1	57.5	1.73E+04	131
Jul	734	-3.4	84.2	2.64E+04	163
Ago	560	0.7	57.8	1.27E+04	112
Sep	776	-0.2	53.9	1.47E+04	121
Oct	370	-0.8	25.1	3.60E+03	60
Nov	480	1.2	23.5	3.65E+03	60
Dic	296	0.8	10.5	9.66E+02	31

## 9.4 Modelo de NAR

Bajo la premisa de que el número de rezagos más adecuado se obtiene a partir del análisis de la función de autocorrelación y de autocorrelación parcial, se determinó que el número óptimo de retrasos para un modelo de red neuronal no lineal con entradas autorregresivas, mejor conocido como modelo NAR es de 24 rezagos.

Partiendo de esta base y una vez normalizado los valores (ecuación 9.1), se generaron modelos NAR variando el número de neuronas en la capa oculta, empezando con 5 e incrementando este número de 5 en 5 hasta llegar un máximo de 30 neuronas. Además, se estabilizaron los pesos mediante el reentrenamiento de la red hasta obtener el modelo que, de acuerdo a las pruebas de desempeño se ajustó mejor a los datos de entrenamiento.

Los resultados de muestran en la **Tabla 9. 7** la cual nos indica que el número de neuronas ocultas óptimo, el cual se obtiene con 25 neuronas. Por lo que el modelo final es un modelo NAR con 25 neuronas en la capa oculta y 24 retrasos.

### 9.4.1 Normalización de los datos de entrada.

Utilizando la ecuación 9.1 se normalizan los valores de entrada con el propósito de disminuir el tiempo de entrenamiento y la sobresaturación de la neurona [11][81]:

$$y = y_{\min} + \frac{x - x_{\min}}{x_{\max} - x_{\min}} (y_{\max} - y_{\min}) \quad (9.1)$$

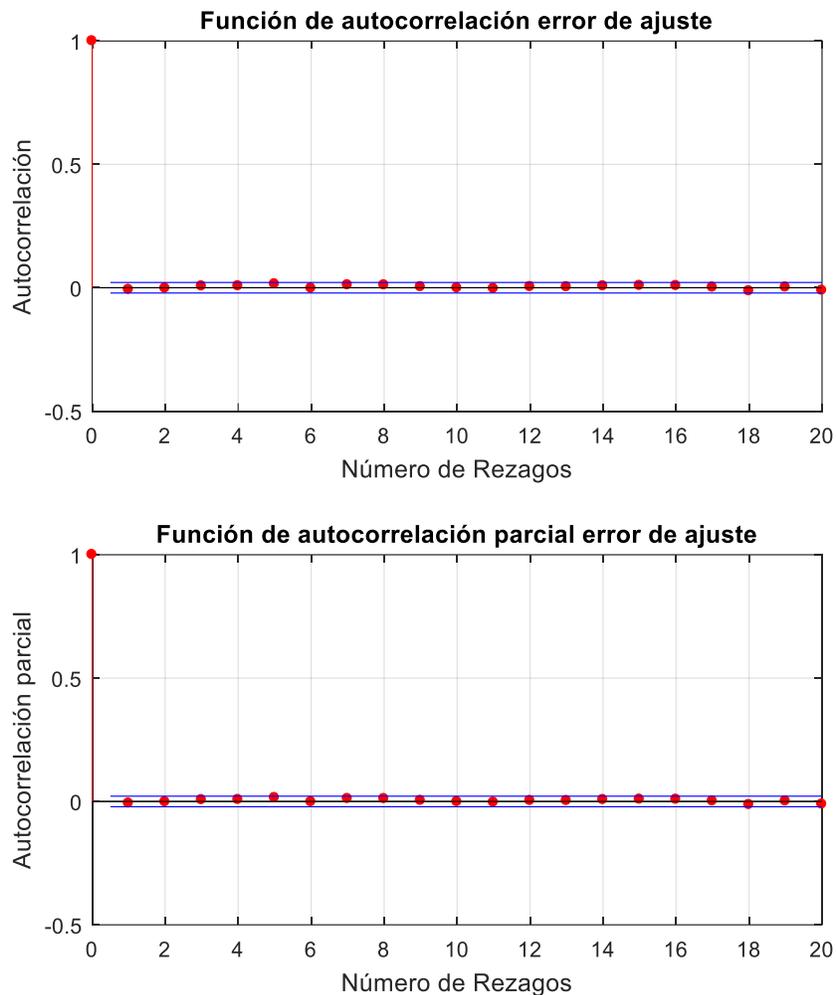
Dónde  $x \in [x_{\min}, x_{\max}]$  y  $y \in [y_{\min}, y_{\max}]$ ;  $x$  es el valor original de los datos y  $y$  es la variable normalizada. Finalmente, en nuestro caso  $y_{\min} = 0$  y  $y_{\max} = 1$  son los valores que han sido asumidos.

**Tabla 9. 7.** Pruebas de modelo NAR utilizando diferente número de neuronas

Neuronas	performance	trainPerformance	valPerformance	testPerformance	Tiempo
5	2.40E-03	2.36E-03	2.36E-03	2.64E-03	00:00:03
10	2.32E-03	2.06E-03	2.93E-03	2.89E-03	00:00:05
15	2.35E-03	2.17E-03	3.10E-03	2.45E-03	00:00:04
20	2.47E-03	2.44E-03	2.53E-03	2.53E-03	00:00:05
25	2.28E-03	2.28E-03	2.33E-03	2.25E-03	00:00:08
30	2.22E-03	2.12E-03	2.24E-03	2.63E-03	00:00:10
				Tiempo Total	00:00:35

#### 9.4.2 Análisis y monitoreo de los errores modelo NAR

En la **Figura 9. 5** se muestra la función de autocorrelación y autocorrelación parcial de los errores.



**Figura 9. 5.** Función de autocorrelación y autocorrelación parcial de los errores, Modelo NAR

De acuerdo a la función de autocorrelación y autocorrelación parcial, los residuos son ruido blanco, por lo que se puede concluir que el modelo NAR es adecuado para ser utilizado en la predicción

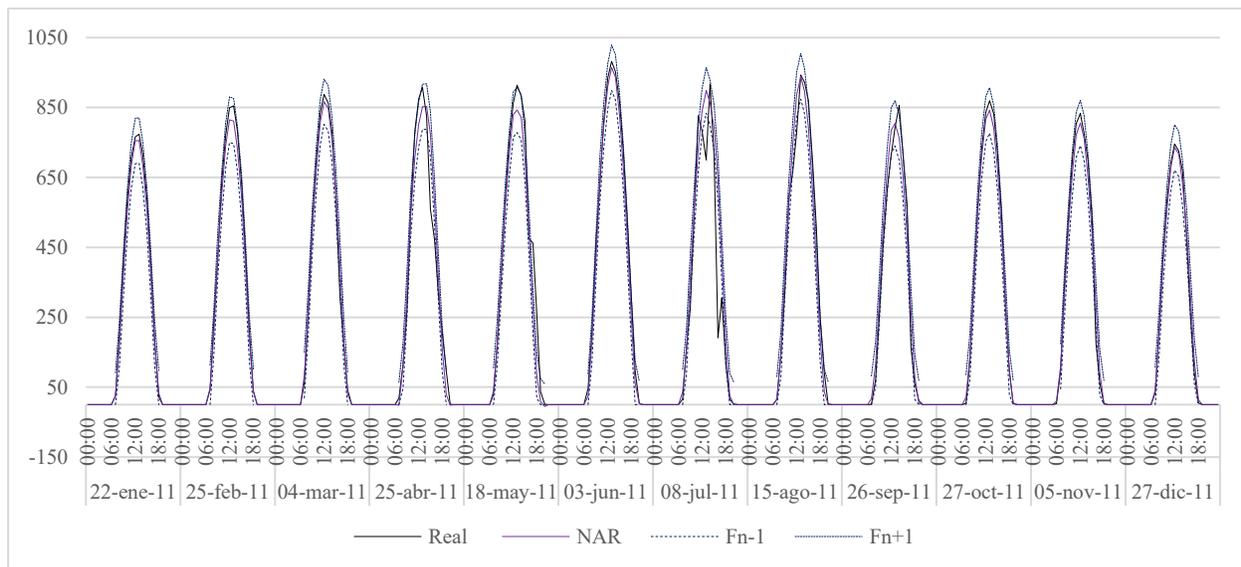
de series de tiempo. La **Tabla 9. 8** muestra los resultados de las pruebas de desempeño para la curva de ajuste, el pronóstico del año 2011 y un día de cada uno de los meses del año 2011.

**Tabla 9. 8.** Resultado NAR de las pruebas de desempeño, Curva de ajuste, Datos predichos, Un día de cada mes

	Precisión	ME	MSE	RMSE	MAE
<b>Curva de ajuste 2010</b>	469.7	-0.203	24.73	2.55E+03	50.50
<b>Pronóstico 2011</b>	722.9	-3.838	32.67	5.37E+03	73.25
<b>Un día para cada mes</b>	417.9	2.036	20.99	2.30E+03	47.98

### 9.4.3 Implementación del modelo NAR para pronóstico de 24 horas adelante del año 2011

En la **Figura 9. 6** se representa el resultado del pronóstico generado mediante el modelo NAR, el pronóstico realizado con este modelo, se elaboro para los mismos días pronosticados con el modelo ARIMA estacional, concretamente, los días 22 de enero, 25 de febrero, 4 de marzo, 25 de abril, 18 de mayo, 3 de junio, 8 de julio, 15 de agosto, 26 de septiembre, 27 de octubre, 5 de noviembre y 27 de diciembre. Esto con el propósito de comparación entre ambos modelos.



**Figura 9. 6.** Pronóstico modelo NAR para 1 día de cada mes del año 2011 elegido de forma aleatoria

Haciendo una comparación de la **Tabla 9. 4** contra la **Tabla 9. 8** notamos que el modelo NAR fue ligeramente superior al modelo ARIMA estacional.

De la misma manera que en el caso del modelo ARIMA se elaboró un reporte de los resultados de las pruebas de desempeño para un día de cada mes y un reporte para cada uno de los meses de año 2011, los resultados se proporcionan en las **Tabla 9. 9** y **Tabla 9. 10**, respectivamente.

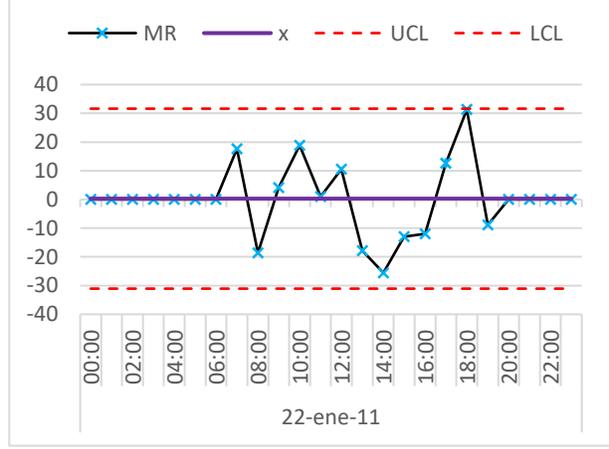
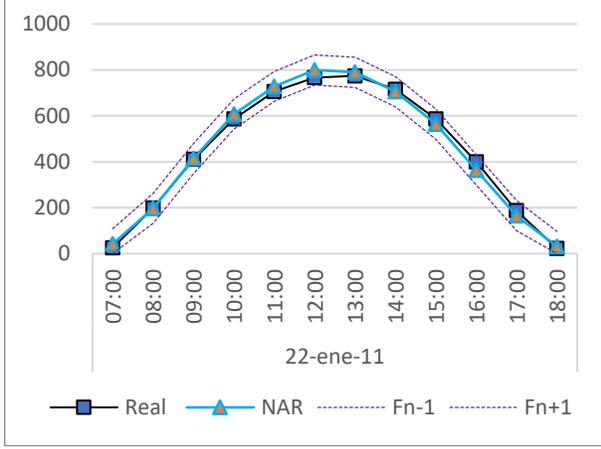
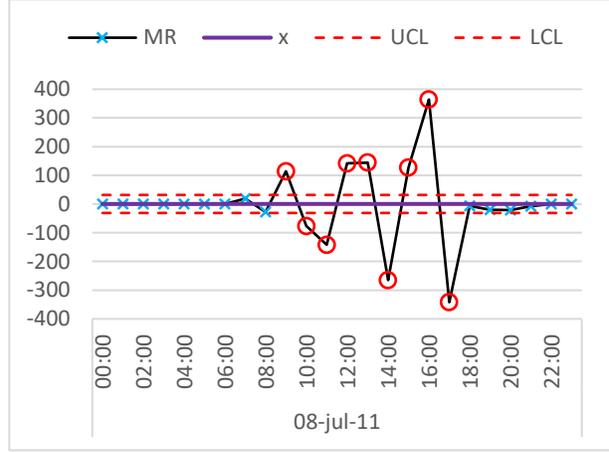
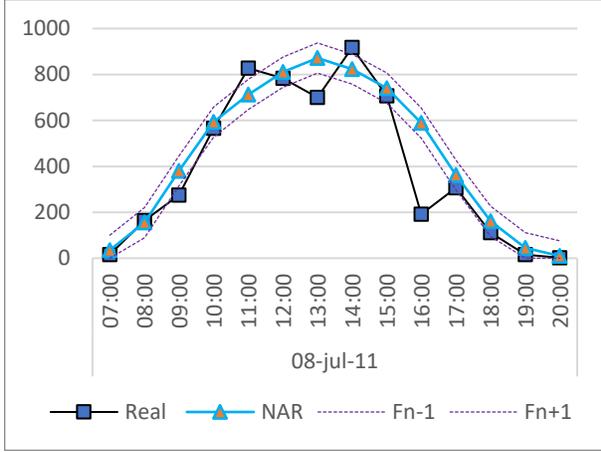
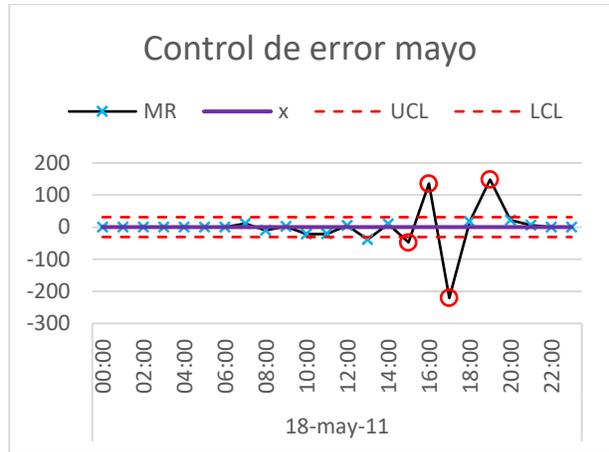
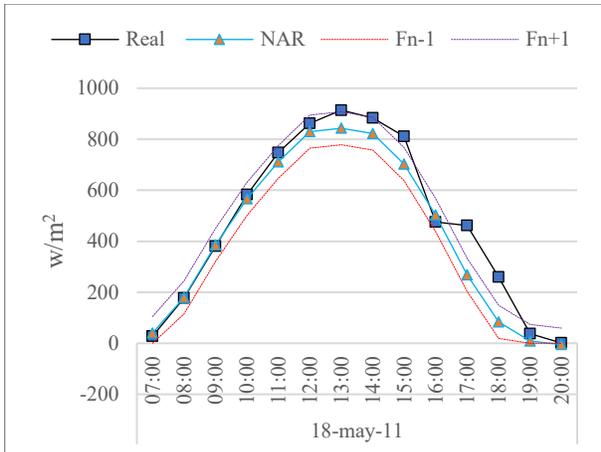
**Tabla 9. 9.** Pruebas de desempeño individuales para un día de cada mes

Mes	Precisión	ME	MAE	MSE	RMSE
22-ene-11	36.5	-10.3	11.3	3.15E+02	17.8
25-feb-11	50.7	-12.9	14.0	5.05E+02	22.5
04-mar-11	96.0	-0.4	11.4	5.31E+02	23.0
25-abr-11	213.5	3.7	33.0	3.48E+03	59.0
18-may-11	193.7	-28.6	32.4	3.90E+03	62.5
03-jun-11	62.4	-15.7	16.9	6.52E+02	25.5
08-jul-11	453.8	39.0	52.3	1.23E+04	110.7
15-ago-11	108.7	-7.7	28.0	2.20E+03	46.9
26-sep-11	103.2	6.8	27.8	1.93E+03	43.9
27-oct-11	45.6	-7.7	10.9	3.25E+02	18.0
05-nov-11	106.8	-3.9	17.1	8.85E+02	29.8
27-dic-11	59.3	0.7	10.2	3.22E+02	18.0

**Tabla 9. 10.** Resultados NAR de las pruebas de desempeño para cada mes de pronóstico

Mes	Precisión	ME	MAE	MSE	RMSE
Ene	207	-4.7	15.3	1.04E+03	32
Feb	311	-8.6	17.3	1.44E+03	38
Mar	225	-11.5	19.1	1.30E+03	36
Abr	511	-7.1	25.7	2.75E+03	52
May	364	-5.3	27.5	2.94E+03	54
Jun	723	4.6	51.3	1.17E+04	108
Jul	585	2.0	69.1	1.66E+04	129
Ago	451	-1.7	49.0	8.63E+03	93
Sep	643	2.7	53.0	1.13E+04	106
Oct	344	-8.1	29.0	3.38E+03	58
Nov	335	-4.4	22.0	2.22E+03	47
Dic	296	-3.9	12.9	9.83E+02	31

También proporcionamos la gráfica de radiación solar real versus la pronosticada con sus respectivos intervalos de predicción, tanto para los dos mejores pronósticos (22 de enero y 27 de octubre), como para los dos peores resultados (8 de julio y 15 de agosto), estos están acompañados de su respectiva caja de control, la cual se utiliza para el monitoreo del error de pronóstico.



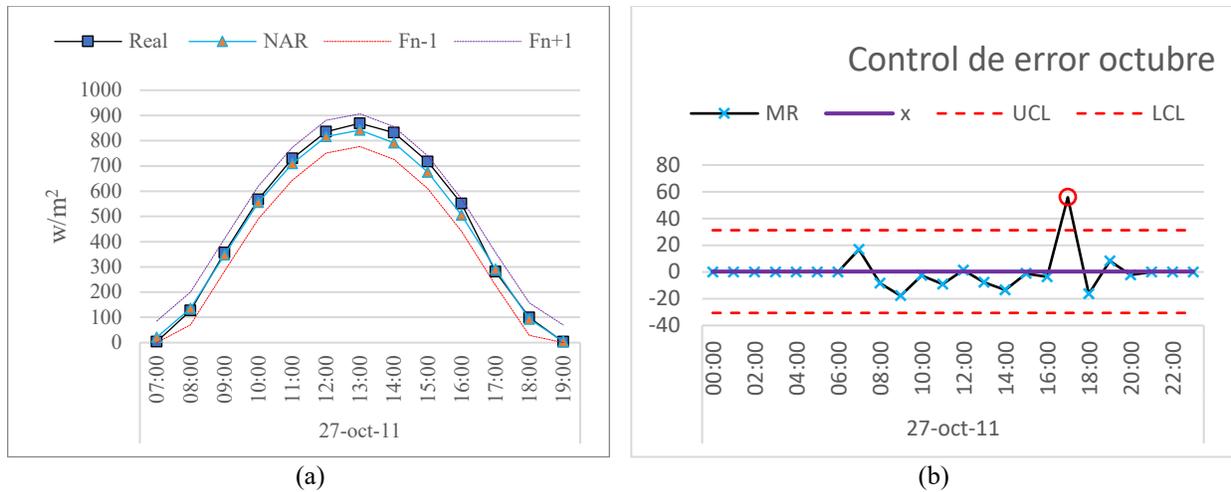


Figura 9. 7. Columna (a) Resultados del pronóstico, Columna (b) caja de control. Modelo NAR

### 9.5 Modelo NARX

Siguiendo el mismo razonamiento aplicado en los dos casos anteriores, el número de rezagos para el modelo NARX se configura en 24, el cual se obtuvo a partir del análisis de la función de autocorrelación y autocorrelación parcial (sección 6.2). Como se mencionó en la (sección 7.7), el modelo NARX es un modelo similar al modelo NAR, que permite el uso de entradas externas con la finalidad de mejorar la estimación de la salida.

#### 9.5.1 Determinación de las entradas del modelo NARX

Las entradas en el molo NARX se determinaron a partir de pruebas de cointegración, específicamente la prueba de causalidad Engle – Granger, sección 5.5.1. Esta prueba funciona solamente con series de tiempo que son estacionarias, es decir media constante y varianza finita, por lo que antes de aplicar la prueba, las series de tiempo son sometidas al test de raíz unitaria Dickey – Fuller, el cual se explica en la sección 5.3, una vez que se ha comprobado que las series de tiempo son estacionarias, se generan los vectores autorregresivos y se aplican las pruebas de causalidad, en la **Tabla 9. 11**, se muestran las variables meteorológicas que se tienen registradas, dónde de acuerdo a la prueba de causalidad, las variables de entrada que se utilizan en el modelo NARX son Temperatura del aire, Humedad relativa y presión.

Tabla 9. 11. Variables meteorológicas Temixco

Sitio	Años de medición	Irradiación global ( $G_I$ )	Temperatura del aire ( $T_a$ )	Humedad relativa ( $RH$ )	Presión ( $P$ )	Velocidad del viento ( $WS$ )
Temixco	2010,2011	*	*	*	*	*

### 9.5.2 Normalización de los datos

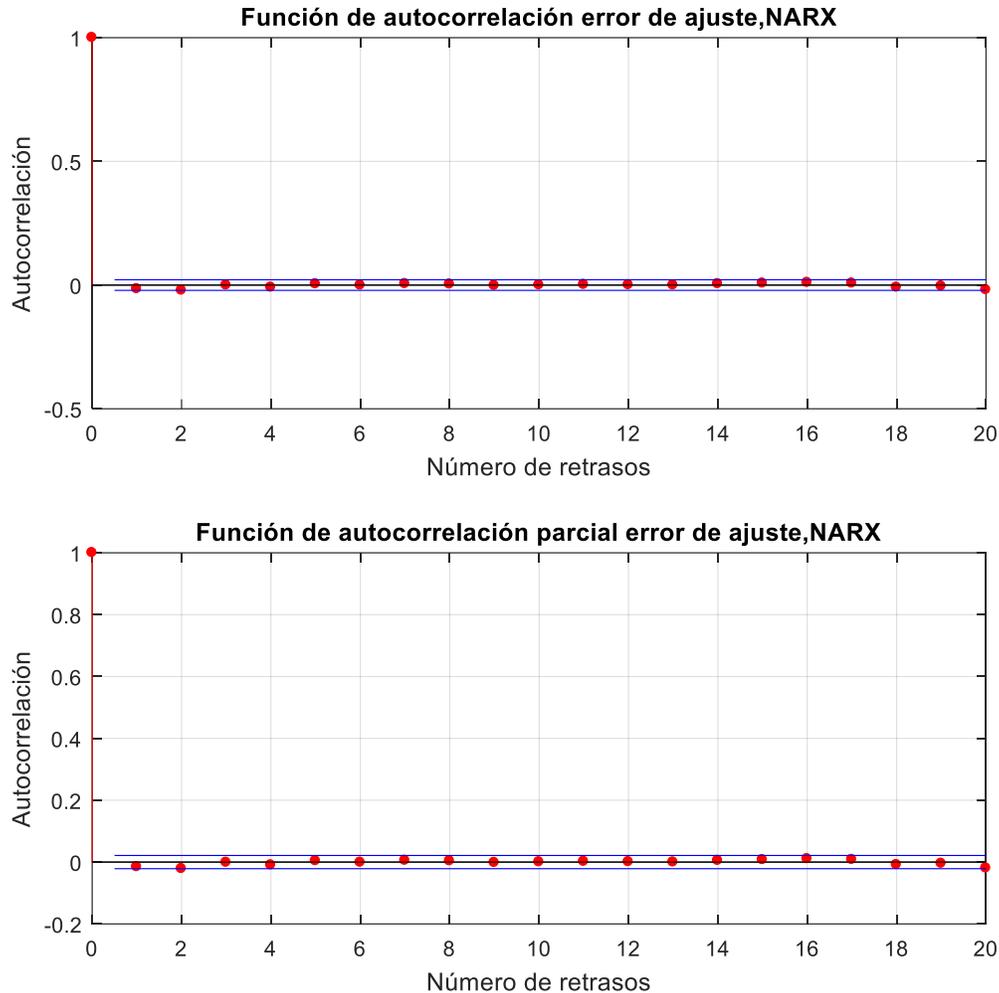
Como regla general y como se había comentado anteriormente, los datos deben de ser normalizados para poder ser utilizados en un modelo NARX. Las variables meteorológicas se normalizaron utilizando la [ecuación 9.1](#), la misma con la que se normalizo la radiación solar en el caso del modelo univariable NAR. En este caso, los mejores resultados se obtuvieron con valores de  $y_{\min} = -1$  y  $y_{\max} = 1$ .

**Tabla 9. 12.** Pruebas del modelo NARX utilizando diferente número de neuronas

Neuronas	performance	trainPerformance	valPerformance	testPerformance	Tiempo
5	8.43E-03	7.76E-03	1.03E-02	9.65E-03	00:01:05
10	8.40E-03	8.58E-03	8.51E-03	7.63E-03	00:00:28
15	8.43E-03	8.13E-03	9.06E-03	9.20E-03	00:00:28
20	8.43E-03	8.64E-03	7.94E-03	7.93E-03	00:00:28
25	8.43E-03	8.31E-03	8.01E-03	9.41E-03	00:00:28
30	8.43E-03	8.66E-03	8.07E-03	7.70E-03	00:00:28
Tiempo Total					00:03:25

En la **Tabla 9. 12** se muestran los resultados del modelo NARX para diferente número de neuronas, el mejor resultado corresponde al modelo para el cual se utilizaron 10 neuronas, como podemos ver los tiempos de calculo so mucho mayores a los tiempos alcanzados por los modelos NAR, esto se debe a la complejidad del modelo ARIMA. De la misma manera las pruebas de desempeño resultan mucho mayores a las del modelo NAR, esto se debe, a que en este modelo los valores se normalizaron en el rango de -1 a 1.

En la Figura se muestran los resultados de la función de autocorrelación y autocorrelación parcial las cuales se aplicaron a los residuos de la curva de ajuste del modelo NARX, como podemos ver los resultados indican que los residuos tienen un comportamiento muy similar al ruido blanco, por lo que este análisis visual nos indica que el modelo es adecuado para el pronóstico.

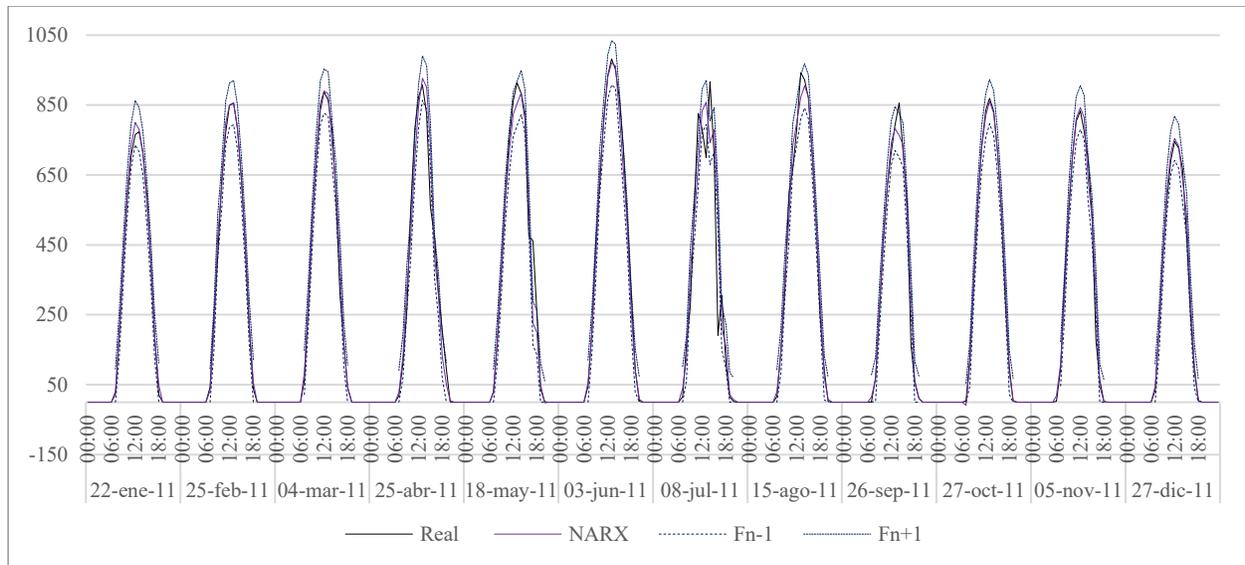


**Figura 9. 8.** Función de autocorrelación y autocorrelación parcial de los errores, Modelo NARX

### 9.5.3 Implementación del modelo NARX para pronóstico de 24 horas adelante del año 2011

Siendo seleccionado la mejor red neuronal se procede a utilizarla para generar el pronóstico 24 horas adelante del año 2011. En la figura se muestra el resultado de 12 días de pronóstico, un día para cada mes del año, estos días corresponden a los reportados en los modelos anteriores, determinación tomada con fines de comparativos.

La **Tabla 9. 13** y **Tabla 9. 14** representan las principales pruebas de desempeño para un día de cada mes y para cada mes, respectivamente. Como podemos ver el modelo NARX supero ampliamente a los modelos univariables, es decir al modelo ARIMA y al modelo NAR, esto se debe a que este modelo no solo toma en cuenta el efecto de sus valores pasados, sino que también considera las perturbaciones generadas por variable exógenas.



**Figura 9. 9.** Pronóstico modelo NARX para 1 día de cada mes del año 2011 elegido de forma aleatoria

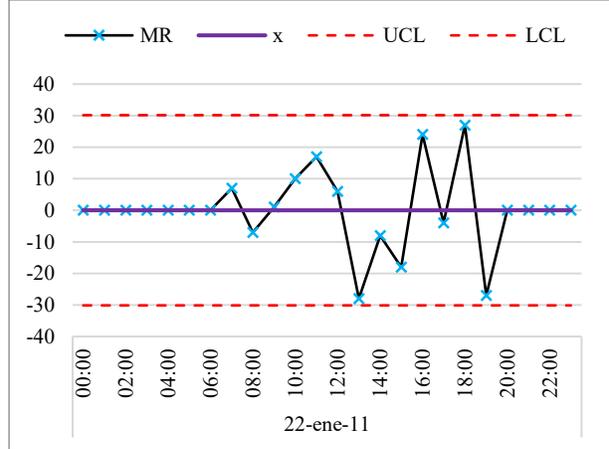
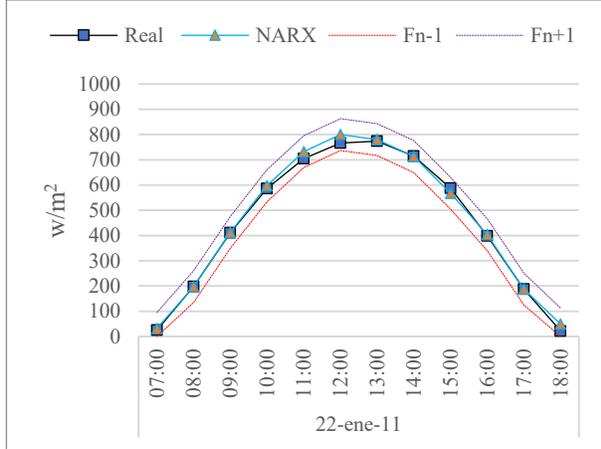
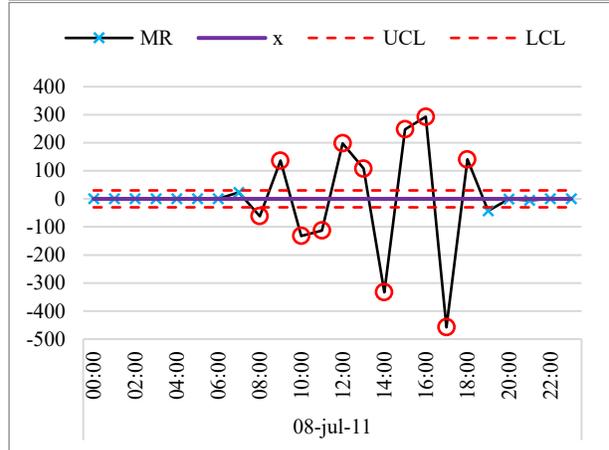
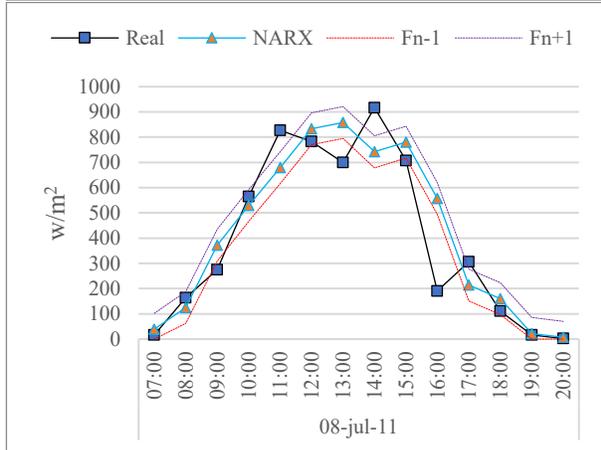
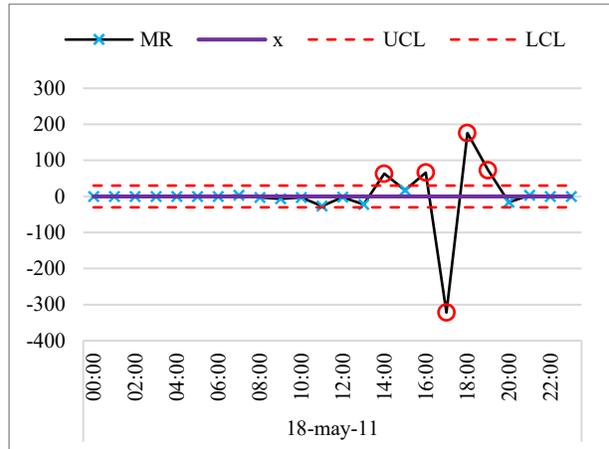
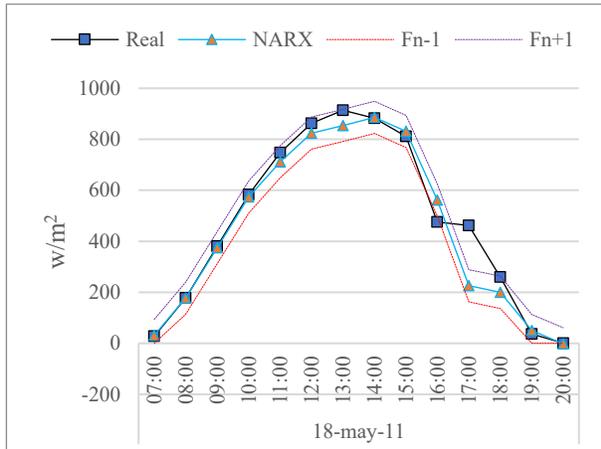
**Tabla 9. 13.** Pruebas de desempeño individuales para un día de cada mes

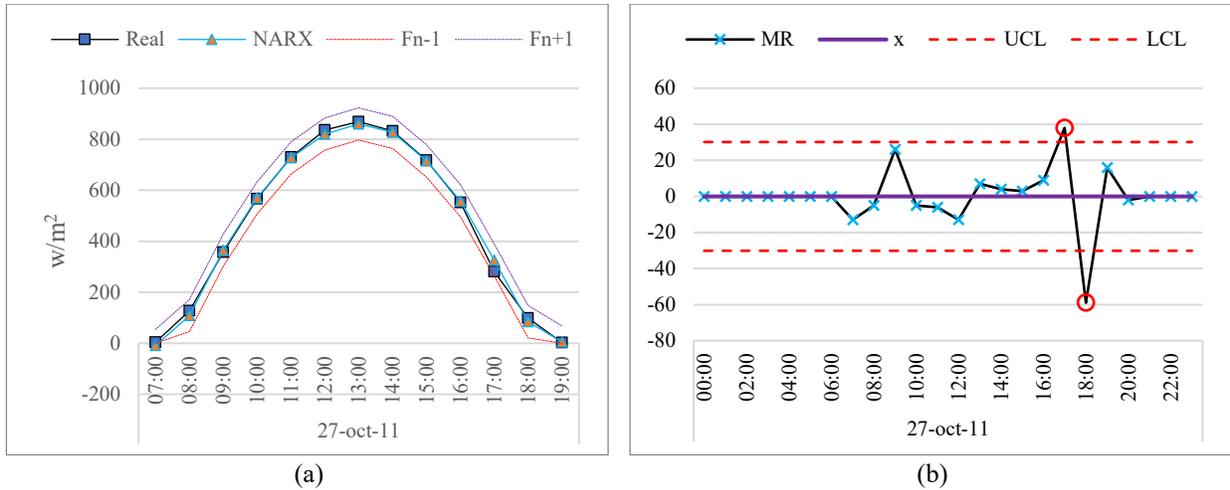
Mes	Precisión	ME	MAE	MSE	RMSE
22-ene-11	34.0	4.0	5.8	1.37E+02	11.7
25-feb-11	54.0	2.5	7.9	2.39E+02	15.5
04-mar-11	97.0	6.4	10.2	5.21E+02	22.8
25-abr-11	177.0	-0.2	30.9	2.78E+03	52.7
18-may-11	236.0	-13.5	23.9	3.07E+03	55.4
03-jun-11	58.0	-8.0	12.8	4.99E+02	22.3
08-jul-11	366.0	14.1	54.8	1.01E+04	100.6
15-ago-11	72.0	0.2	16.0	8.22E+02	28.7
26-sep-11	152.0	2.7	16.1	1.44E+03	37.9
27-oct-11	45.0	-0.6	6.0	1.34E+02	11.6
05-nov-11	135.0	4.7	14.8	1.07E+03	32.7
27-dic-11	58.0	5.9	9.8	3.27E+02	18.1

**Tabla 9. 14.** Resultados NARX de las pruebas de desempeño para cada mes de pronóstico

Mes	Precisión	ME	MAE	MSE	RMSE
Ene	280	0.0	10.6	5.95E+02	24
Feb	302	0.4	11.1	8.65E+02	29
Mar	280	0.8	11.1	7.41E+02	27
Abr	504	1.6	16.6	1.75E+03	42
May	339	0.6	20.4	2.01E+03	45
Jun	430	0.0	30.5	4.10E+03	64
Jul	464	1.2	40.3	6.19E+03	79
Ago	316	-2.3	32.5	3.92E+03	63

Sep	432	0.3	30.1	3.92E+03	63
Oct	298	-1.0	18.8	1.72E+03	41
Nov	189	1.8	11.9	6.31E+02	25
Dic	182	0.0	7.8	2.93E+02	17





**Figura 9. 10.** Columna (a) Resultados del pronóstico, Columna (b) caja de control. Modelo NARX

En la **Figura 9. /Figura 9. 10** se presenta el detalle de los peores pronósticos y de los mejores dos pronósticos, en la columna (a), hacemos la comparación del valor medido contra el valor pronosticado, además se ha incluido los intervalos de predicción, mientras que en la columna (b), se muestra la caja de control correspondiente a cada uno de los pronósticos.

## Conclusiones.

Este trabajo de investigación consta de dos partes principales, una que trata con el desarrollo de modelos de predicción de la radiación solar, y la otra que trata con la generación de un pronóstico de la potencia generada por un sistema fotovoltaico a partir del pronóstico de la radiación solar. Para lograr el desarrollo del mismo, fue necesario obtener variables meteorológicas del sitio de interés, analizar dichas variables mediante la implementación de técnicas estadísticas y métodos de análisis y transformación de series de tiempo, posteriormente se generaron tres modelos de predicción, ARIMA estacional, NAR y NARX, estos fueron comparados entre sí, llegando a la conclusión de que los modelos de redes neuronales son superiores a las técnicas tradicionales ARIMA. Finalmente, se obtuvo la potencia de un sistema fotovoltaico mediante los resultados de pronóstico de la radiación solar.

## Referencias Biográficas.

- [1] A. Sfetsos, A.H. Coonick, Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques, *Sol. Energy*. 68 (2000) 169–178. doi:10.1016/S0038-092X(99)00064-X.
- [2] J. Cao, X. Lin, Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks, *Energy Convers. Manag.* 49 (2008) 1396–1406. doi:10.1016/j.enconman.2007.12.030.
- [3] F.O. Hocaoglu, Ö.N. Gerek, M. Kurban, Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed-forward neural networks, *Sol. Energy*. 82 (2008) 714–726. doi:10.1016/j.solener.2008.02.003.
- [4] N.S. Maimouna Diagne, Mathieu David, Philippe Lauret, John Boland, Review of solar irradiance forecasting methods and proposition for small-scale insular grids, *Renew. Sustain. Energy Rev.* (2013).
- [5] M. a. Behrang, E. Assareh, a. Ghanbarzadeh, a. R. Noghrehabadi, The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data, *Sol. Energy*. 84 (2010) 1468–1480. doi:10.1016/j.solener.2010.05.009.
- [6] C. Chen, S. Duan, T. Cai, B. Liu, Online 24-h solar power forecasting based on weather type classification using artificial neural network, *Sol. Energy*. 85 (2011) 2856–2870. doi:10.1016/j.solener.2011.08.027.
- [7] K.S. Reddy, M. Ranjan, Solar resource estimation using artificial neural networks and comparison with other correlation models, *Energy Convers. Manag.* 44 (2003) 2519–2530. doi:10.1016/S0196-8904(03)00009-8.
- [8] A.S.S. Dorvlo, J. a. Jervase, A. Al-Lawati, Solar radiation estimation using artificial neural networks, *Appl. Energy*. 71 (2002) 307–319. doi:10.1016/S0306-2619(02)00016-8.
- [9] A. Sözen, E. Arcaklioglu, M. Özalp, N. Çağlar, Forecasting based on neural network approach of solar potential in Turkey, *Renew. Energy*. 30 (2005) 1075–1090. doi:10.1016/j.renene.2004.09.020.
- [10] J. Mubiru, E.J.K.B. Banda, Estimation of monthly average daily global solar irradiation using artificial neural networks, *Sol. Energy*. 82 (2008) 181–187. doi:10.1016/j.solener.2007.06.003.
- [11] A. Mellit, A.M. Pavan, A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy, *Sol. Energy*. 84 (2010) 807–821. doi:10.1016/j.solener.2010.02.006.
- [12] O. Şenkal, T. Kuleli, Estimation of solar radiation over Turkey using artificial neural network and satellite data, *Appl. Energy*. 86 (2009) 1222–1228. doi:10.1016/j.apenergy.2008.06.003.
- [13] A. Mellit, A.H. Arab, N. Khorissi, H. Salhi, An ANFIS-based forecasting for solar radiation data from sunshine duration and ambient temperature, (2007). <http://dx.doi.org/10.1109/PES.2007.386131>.
- [14] M. Mellit, S. Shaari, H. Mekki, N. Khorissi, FPGA-based artificial neural network for prediction of solar radiation data from sunshine duration and air temperature, 2008 IEEE Reg. 8 Int. Conf. Comput. Technol. Electr. Electron. Eng. (2008) 118–123. doi:10.1109/SIBIRCON.2008.4602597.
- [15] a. Ghanbarzadeh, a. R. Noghrehabadi, E. Assareh, M. a. Behrang, Solar radiation forecasting based on meteorological data using artificial neural networks, 2009 7th IEEE Int. Conf. Ind. Informatics. (2009) 227–231. doi:10.1109/INDIN.2009.5195808.
- [16] F. Deng, C. Liu, Z. Wang, Global Solar Radiation Modeling Using The Artificial Neural Network Technique, (2010).
- [17] F. Nomiya, J. Asai, T. Murakami, H. Takano, J. Murata, A study on global solar radiation forecasting models using meteorological data and their application to wide area forecast, 2012 IEEE Int. Conf. Power Syst. Technol. (2012) 1–6. doi:10.1109/PowerCon.2012.6401336.
- [18] P. Zhang, H. Takano, J. Murata, Daily solar radiation prediction based on wavelet analysis, SICE Annu. Conf. 2011. (2011) 712–717.
- [19] H. Mori, a. Takahashi, A data mining method for selecting input variables for forecasting model of global solar radiation, *Pes T&D 2012*. (2012) 1–6. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6281569>.
- [20] R. Huang, T. Huang, R. Gadh, N. Li, Solar generation prediction using the ARMA model in a laboratory-level micro-grid, 2012 IEEE 3rd Int. Conf. Smart Grid Commun. SmartGridComm 2012. (2012) 528–533. doi:10.1109/SmartGridComm.2012.6486039.
- [21] N. Zhang, C. Williams, P.K. Behera, Solar radiation prediction based on particle swarm optimization and evolutionary algorithm using recurrent neural networks, *IEEE*. (2013).
- [22] E. Akarslan, F.O. Hocaoglu, A novel adaptive approach for hourly solar radiation forecasting, *Renew. Energy*. 87 (2016) 628–633. doi:10.1016/j.renene.2015.10.063.
- [23] J. Boland, M. David, P. Lauret, Short term solar radiation forecasting: Island versus continental sites, *Energy*. 113 (2016) 186–192. doi:10.1016/j.energy.2016.06.139.
- [24] P.F. Jiménez-Pérez, L. Mora-López, Modeling and forecasting hourly global solar radiation using clustering and classification techniques, *Sol. Energy*. 135 (2016) 682–691. doi:10.1016/j.solener.2016.06.039.
- [25] L. Martín, L.F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, M. Cony, Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning, *Sol. Energy*. 84 (2010) 1772–1781. doi:10.1016/j.solener.2010.07.002.
- [26] S. Monjoly, M. André, R. Calif, T. Soubdhan, Hourly forecasting of global solar radiation based on multiscale

- decomposition methods: A hybrid approach, *Energy*. 119 (2017) 288–298. doi:10.1016/j.energy.2016.11.061.
- [27] R. Azimi, M. Ghayekhloo, M. Ghofrani, A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting, *Energy Convers. Manag.* 118 (2016) 331–344. doi:10.1016/j.enconman.2016.04.009.
- [28] K. Benmouiza, A. Chekneane, Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models, *Energy Convers. Manag.* 75 (2013) 561–569. doi:10.1016/j.enconman.2013.07.003.
- [29] C.-R. Chen, U. Kartini, k-Nearest Neighbor Neural Network Models for Very Short-Term Global Solar Irradiance Forecasting Based on Meteorological Data, *Energies*. 10 (2017) 186. doi:10.3390/en10020186.
- [30] C. Voyant, M. Muselli, C. Paoli, M.L. Nivet, Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation, *Energy*. 39 (2012) 341–355. doi:10.1016/j.energy.2012.01.006.
- [31] C. Voyant, M. Muselli, C. Paoli, M.-L. Nivet, Hybrid methodology for hourly global radiation forecasting in Mediterranean area, *Renew. Energy*. 53 (2013) 1–11. doi:10.1016/j.renene.2012.10.049.
- [32] W. Ji, K.C. Chee, Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN, *Sol. Energy*. 85 (2011) 808–817. doi:10.1016/j.solener.2011.01.013.
- [33] C. Renno, F. Petito, A. Gatto, ANN model for predicting the direct normal irradiance and the global radiation for a solar application to a residential building, *J. Clean. Prod.* 135 (2016) 1298–1316. doi:10.1016/j.jclepro.2016.07.049.
- [34] S. Kaplanis, E. Kaplani, Stochastic prediction of hourly global solar radiation for Patra, Greece, *Appl. Energy*. 87 (2010) 3748–3758. doi:10.1016/j.apenergy.2010.06.006.
- [35] P.K. Pandey, M.L. Soupir, A new method to estimate average hourly global solar radiation on the horizontal surface, *Atmos. Res.* 114–115 (2012) 83–90. doi:10.1016/j.atmosres.2012.05.012.
- [36] D. Yang, P. Jirutitijaroen, W.M. Walsh, Hourly solar irradiance time series forecasting using cloud cover index, *Sol. Energy*. 86 (2012) 3531–3543. doi:10.1016/j.solener.2012.07.029.
- [37] J. Wu, C.K. Chan, Prediction of hourly solar radiation with multi-model framework, *Energy Convers. Manag.* 76 (2013) 347–355. doi:10.1016/j.enconman.2013.07.067.
- [38] J. Huang, M. Korolkiewicz, M. Agrawal, J. Boland, Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model, *Sol. Energy*. 87 (2013) 136–149. doi:10.1016/j.solener.2012.10.012.
- [39] F.O. Hocaoglu, F. Karanfil, A time series-based approach for renewable energy modeling, *Renew. Sustain. Energy Rev.* 28 (2013) 204–214. doi:10.1016/j.rser.2013.07.054.
- [40] A. Ahmad, T.N. Anderson, T.T. Lie, Hourly global solar irradiation forecasting for New Zealand, *Sol. Energy*. 122 (2015) 1398–1408. doi:10.1016/j.solener.2015.10.055.
- [41] L.M. Aguiar, B. Pereira, P. Lauret, F. Díaz, M. David, Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting, *Renew. Energy*. 97 (2016) 599–610. doi:10.1016/j.renene.2016.06.018.
- [42] M.C. Di Piazza, G. Vitale, *Photovoltaic Sources: Modeling and Emulation*, 1st ed., Springer London, London, 2013. doi:10.1007/978-1-4471-4378-9.
- [43] J. Duffie, W. Beckman, *Solar Engineering of Thermal Processes*, 3th ed., 2006. doi:10.1115/1.2930068.
- [44] O. Perpiñán Lamigueiro, *Energía Solar Fotovoltaica*, (2013) 192.
- [45] N. Cullen, J. Thornycroft, A. Collinson, Risk analysis of islanding of photovoltaic power systems within low voltage distribution networks, (2002). [http://iea-pvps.org/index.php?id=9&eID=dam\\_frontend\\_push&docID=385](http://iea-pvps.org/index.php?id=9&eID=dam_frontend_push&docID=385).
- [46] J.J. Michalsky, The Astronomical Almanac's algorithm for approximate solar position (1950-2050), *Sol. Energy*. 40 (1988) 227–235. doi:10.1016/0038-092X(88)90045-X.
- [47] J.W. SPENCER, Fourier Series Representation of the Position of the Sun, *Search*. 2 (1971) 172–172. <http://www.mail-archive.com/sundial@uni-koeln.de/msg01050.html>.
- [48] L. Strous, Position of the Sun, (2011). <http://aa.quae.nl/en/reken/zonpositie.html>.
- [49] G.N. Tiwari, A. Tiwari, Shyam, *Handbook of Solar Energy*, 1st ed., Springer Singapore, 2016. doi:10.1007/978-981-10-0807-8.
- [50] M.E. Mackay, *Solar Energy and Introduction*, Oxford, Oxford, 2015.
- [51] D. Guasch Murillo, *Modelado y análisis de sistemas fotovoltaicos*, (2003) 223.
- [52] E.B. Babatunde, *Solar Radiation*, Janeza Trdine 9, 51000 Rijeka, Croatia, Rijeka, 2012.
- [53] F. Kasten, A.T. Young, Revised optical air mass tables and approximation formula, *Appl. Opt.* 28 (1989) 4735–4738.
- [54] O. Behar, A. Khellaf, K. Mohammedi, Comparison of solar radiation models and their validation under Algerian climate - The case of direct irradiance, *Energy Convers. Manag.* 98 (2015) 236–251. doi:10.1016/j.enconman.2015.03.067.
- [55] B.Y.H. Liu, R.C. Jordan, The interrelationship and characteristic distribution of direct, diffuse and total solar radiation, *Sol. Energy*. 4 (1960) 1–19. doi:10.1016/0038-092X(60)90062-1.
- [56] R. Aguiar, M. Collares-Pereira, Statistical properties of hourly global radiation, *Sol. Energy*. 48 (1992) 157–167.

- doi:10.1016/0038-092X(92)90134-V.
- [57] Ministerio de Ciencia y Tecnología, Real Decreto 842/2002, de 2 de agosto, por el que se aprueba el Reglamento electrotécnico para baja tensión, Boletín Of. Del Estado. (2002) 33084–33086. <https://www.boe.es/boe/dias/2002/09/18/pdfs/A33084-33086.pdf>.
- [58] M. Córdova-Zamora, Estadística descriptiva e inferencial, 1995. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Estadística+descriptiva+e+inferencial#2>.
- [59] M.F. Triola, Estadística, 2004. doi:10.1017/CBO9781107415324.004.
- [60] R.A. Maronna, Probabilidad y Estadística Elementales., La Plata. (1995) 117.
- [61] W. Mendenhall, R. Beaver, B. Beaver, Introducción a la probabilidad y la estadística, 2010. [http://investigadores.cide.edu/aparicio/data/refs/Mendenhall\\_Prob\\_Estadística\\_13.pdf](http://investigadores.cide.edu/aparicio/data/refs/Mendenhall_Prob_Estadística_13.pdf).
- [62] A. Hayter, Probability and statistics for engineers and scientists, 2012. doi:10.1007/s13398-014-0173-7.2.
- [63] G.D. N., D.C. Porter, Basic Econometrics, 5th ed., Douglas Reiner, 2009.
- [64] D. Montgomery, Introduction to statistical quality control, 2009. doi:10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.
- [65] B.H. Baltagi, Econometrics, Fourth, Springer, Berlin Heidelberg, 2008.
- [66] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis: Forecasting & Control, 1994. doi:10.1016/j.ijforecast.2004.02.001.
- [67] J.E. Hanke, D.W. Wichern, Pronosticos en los negocios, 2006. [http://books.google.com.co/books?id=WaiOrL8oct4C&printsec=frontcover&hl=es&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](http://books.google.com.co/books?id=WaiOrL8oct4C&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false).
- [68] D.C. (Arizona S.U. Montgomery, C.L. (Bank of A. Jennings, M. (Technical U. of D. Kulahci, Introduction to Time Series Analysis and Forecasting, Second, Wiley, New Jersey, 2015. doi:10.1007/978-3-319-28725-6.
- [69] S. Makridakis, S.C. Wheelwright, R.J. Hyndman, Forecasting Methods and Applications, John Wiley & Sons, Inc., 1998.
- [70] E. Cadenas, W. Rivera, R. Campos-Amezcu, R. Cadenas, Wind speed forecasting using the NARX model, case: La Mata, Oaxaca, México, Neural Comput. Appl. (2015). doi:10.1007/s00521-015-2012-y.
- [71] J.R.H. Gonzáles, Víctor José Martínez Hernando, Redes neuronales artificiales fundamentos, modelos y aplicaciones, Alfa Omega, México D.F., 2000.
- [72] T. Kohonen, An introduction to neural computing, Neural Networks. 1 (1988) 3–16. doi:10.1016/0893-6080(88)90020-2.
- [73] N. Hecht, Neurocomputing: Picking the human brain, IEEE Spectr. (1988) 36–41.
- [74] S. Haykin, Neural Networks a comprehensive foundation, Second, PEARSON PRENTICE HALL, 2004. doi:0-13-273350-1.
- [75] V.Á. Salas, E. Oli, A. Barrado, A. La, Review of the maximum power point tracking algorithms for stand-alone photovoltaic systems, 90 (2006) 1555–1578. doi:10.1016/j.solmat.2005.10.023.
- [76] E. Skoplaki, J.A. Palyvos, On the temperature dependence of photovoltaic module electrical performance : A review of efficiency / power correlations, Sol. Energy. 83 (2009) 614–624. doi:10.1016/j.solener.2008.10.008.
- [77] A. Yazdani, A. Rita, D. Fazio, H. Ghoddami, M. Russo, M. Kazerani, J. Jatskevich, K. Strunz, S. Leva, J.A. Martinez, Modeling Guidelines and a Benchmark for Power System Simulation Studies of Three-Phase Single-Stage Photovoltaic Systems, IEEE Trans. POWER Deliv. 26 (2011) 1247–1264.
- [78] C. Vimalarani, N. Kamaraj, Modeling and performance analysis of the solar photovoltaic cell model using Embedded MATLAB, (2015). doi:10.1177/0037549714568535.
- [79] M.G. Villalva, J.R. Gazoli, E.R. Filho, MODELING AND CIRCUIT-BASED SIMULATION OF PHOTOVOLTAIC ARRAYS, (2009) 1244–1254.
- [80] L. (Universidad P. de C. Castañer, S. (Universidad P. de C. Silvestre, Modelling Photovoltaic Systems Using PSpice, John Wiley & Sons, LTD, Cataluña, 2002.
- [81] M.H. Beale, M.T. Hagan, H.B. Demuth, Neural Network Toolbox™ User's Guide R 2013 b, (2013).