

Diseño de un modelo inteligente para la predicción horaria a corto plazo de la demanda eléctrica nacional basado en aprendizaje profundo y propuesta de implementación en el despacho de carga de la ANDE

Félix Juan Ramon Morales Mareco^{a,1}, Matías Emmanuel Verdún Bello^{a,1}, Ing. Carlos Sauer^{c,4}, Dr. Diego Stalder^{c,4}

^aAlumno. Facultad de Ingeniería, Universidad Nacional de Asunción, Paraguay

^bTutor. Facultad de Ingeniería, Universidad Nacional de Asunción, Paraguay

^cTutor. Facultad de Ingeniería, Universidad Nacional de Asunción, Paraguay

PALABRAS CLAVE

Consumo Eléctrico
Series Temporales
Predicción
Redes Neuronales
Recurrentes
Algoritmo Genético

RESUMEN

La previsión de la demanda eléctrica a corto plazo es un insumo fundamental para la toma de decisiones por parte de entidades dedicadas a la administración de energía eléctrica. Los modelos predictivos utilizados tradicionalmente (determinísticos y estadísticos) se basan en el análisis de series temporales históricas. Este trabajo propone un modelo inteligente basado en celdas LSTM (*Long-Short Term Memory*), el buque insignia de las Redes Neuronales Recurrentes, cuya optimización de parámetros está a cargo de un Algoritmo Genético para encontrar el mejor desempeño en la predicción horaria del consumo eléctrico. El *dataset* entero abarca datos históricos de la demanda por hora de 2009 a 2018 de la Administración Nacional de Electricidad (ANDE), además de la inclusión de variables climáticas y variables del calendario que se correlacionan con el consumo eléctrico. El modelo desarrollado permite asistir a la ANDE en la toma de decisiones críticas y en la programación diaria de las operaciones de carga en los despachos de las centrales de generación.

KEYWORDS

Electric Consumption
Time Series
Forecasting
Recurrent Neural
Network
Genetic Algorithm

ABSTRACT

The forecast of short-term electricity demand is a fundamental input for decision-making by entities dedicated to the administration of electricity. Traditionally used predictive models (deterministic and statistical) are based on the analysis of historical time series. This paper proposes an intelligent model based on LSTM cells (*Long-Short Term Memory*), the flagship of Recurrent Neural Networks, whose optimization of the parameters are in charge of a Genetic Algorithm to find the best hourly prediction performance. The entire dataset includes historical hourly demand data from 2009 to 2018 from the *Administración Nacional de Asunción* (ANDE), further the inclusion of climatic variables and calendar variables related with electricity consumption. The model would allow ANDE to assist in making critical decisions and in the daily scheduling of electric consume.

1. Introducción

La previsión de la demanda de consumo eléctrico es un insumo fundamental para la toma de decisiones operativas, técnicas y estratégicas por la Administración Nacional de Electricidad (ANDE). Un problema con la energía eléctrica es el alto costo de almacenarlas una vez generada y por tanto, es de vital importancia prever con la mayor precisión posible

la demanda futura. Si la demanda futura predicha es muy superior a la real, puede generar pérdidas económicas por la programación de generación innecesaria, mientras que si la demanda futura predicha es muy inferior, se da lugar a un mayor coste para el consumidor debido a la necesidad de recurrir a generación de contingencia.

La ANDE planifica diaria, mensual y anualmente en distintos escenarios con alta incertidumbre y bajo un amplio margen de factores que pueden afectar a la demanda prevista. Esta planificación es fundamental para poder asegurarse de satisfacer la necesidad del país mediante las fuentes generadoras del país. La literatura sobre modelos de previsión de carga es extensa [13]. Las metodologías tradicionales todavía aplicadas, son aquellas que utilizan técnicas estadísticas, tales como las de regresión lineal y la

¹Alumno de Ingeniería Mecatrónica de la Facultad de Ingeniería de la UNA, Paraguay (email: moralesmfelixjr@gmail.com).

²Alumno de Ingeniería Mecatrónica de la Facultad de Ingeniería de la UNA, Paraguay (email: matias.verdun95@hotmail.com).

³Docente de la Facultad de Ingeniería, Universidad Nacional de Asunción (email: csauer@ing.una.py).

⁴Docente de la Facultad de Ingeniería, Universidad Nacional de Asunción (email: dstalder@ing.una.py).

metodología clásica de Box-Jenkins con el análisis de series temporales [10].

Recientemente, se está prestando especial atención a modelos enfocados en *Deep Learning* (aprendizaje profundo). Estos se basan en redes neuronales artificiales (RNA), con arquitecturas más profundas y complejas [9] que ha despertado interés la comunidad científica debido a su capacidad de capturar el comportamiento de los datos, inclusive cuando se consideran patrones no lineales complejos y grandes cantidades de datos. Existen diferentes modelos y arquitecturas de RNA, las cuales son seleccionadas teniendo de acuerdo con la aplicación de interés. Dentro de este conjunto se encuentran las redes neuronales recurrentes (RNN), las cuales aprovechan un tipo especial de capa neural, conocidas como capas recurrentes con memoria, lo que las hace ideal para predecir series temporales a partir de datos históricos del uso de la red eléctrica [5]. La célula LSTM (Long-Short Term Memory), una arquitectura evolucionada de las RNN, es una opción atractiva para modelar datos secuenciales como series de tiempo, ya que el principio básico detrás de su diseño es que permite incluir otras variables y transmitir información importante de manera confiable en el futuro, gracias a la célula de memoria que posee. Ésta contiene información crítica que ha aprendido con el tiempo, y la red está diseñada para mantener de manera efectiva la información útil en ella, durante muchos pasos de tiempo [7].

2. Breve reseña del sistema eléctrico de Paraguay

El Paraguay es un país con características únicas en lo que se refiere a su desarrollo eléctrico y sus relaciones con el desarrollo socioeconómico. Posee una ventaja comparativa y competitiva crucial en la actualidad: abundancia de energía eléctrica generada a partir del aprovechamiento de fuentes de energías hidroeléctrica [1].

Actualmente el Paraguay posee dos grandes centrales hidroeléctricas binacionales y una propia:

- **Itaipú (CH-IPU):** Compartido con el Brasil e instalado sobre el río Paraná. Posee una capacidad de 14.000 MW, correspondiéndole equitativamente por derecho a ambos países. Sin embargo, Paraguay solo usa una pequeña fracción de la potencia y energía que le corresponde en un largo periodo de tiempo. Por esa razón, tras múltiples notas reversales del contrato [?], actualmente la ANDE tiene prioridad de definir su necesidad de contratación de potencia de su misma parte, y toda aquella no estipulada, la entidad brasileña obligatoriamente debería comprarlo. La *potencia contratada* por parte de la ANDE se hará de forma anual, especificando las potencias requeridas mensualmente, y que la *potencia excedente* disponible debido a la disponibilidad energética sobrantes de las unidades generadoras producida por la eficiencia alcanzadas por las mismas y el aumento del salto bruto [?], será utilizada prioritariamente para solventar el déficit de consumo del ente paraguayo. En manifiesto a esto, toda la planificación diaria que realiza la ANDE para el SIN, gira en torno al conocimiento de la existencia o no de esta potencia excedente. Cuando la misma es insuficiente para cubrir el intercambio ANDE-Itaipú, se recalca en la figura de *cesión de potencia*, por parte de la entidad brasileña a la ANDE, cuyo costo es superior a la de la potencia contratada [3]

- **Yacyretá (CH-YAC):** instalada sobre el río Paraná a la altura de la isla Yacyretá, compartiendo con Argentina. En el año 2011 finalmente pudo llegar a su capacidad máxima de 3200 MW, tras varios acuerdos políticos-económicos entre ambos países. Es conveniente resaltar que la ANDE compra energía de Yacyretá (a diferencia de la contratación de potencia de Itaipú). De esta manera, la programación de la operación o el despacho en Yacyretá, no atrae mayores riesgos económicos en relación a la potencia disponible por la central hidroeléctrica para la ANDE, ya que se paga solo por la energía utilizada, cuyo costo se encuentra en torno a los 40.5 US\$/MWh, de acuerdo a la Nota Reversal del año 1992 (NR92) [?]
- **Acaray(CH-ACY):** La CH-ACY con una capacidad instalada de 214 MW, es total propiedad de la ANDE. Por consiguiente, el costo de la energía generada, puede estimarse considerando los gastos de operación y mantenimiento de la central (que incluyen operación y mantenimiento de maquinarias) y la energía total generada en un año.

2.1. Demanda eléctrica

En contexto al sistema eléctrico nacional, la demanda de potencia ha acompañado la evolución de los requerimientos de energía de la población para el desarrollo de las diversas actividades, acentuándose los picos de máxima potencia por condiciones climáticas eventuales como las altas temperaturas y el elevado nivel de humedad [3].

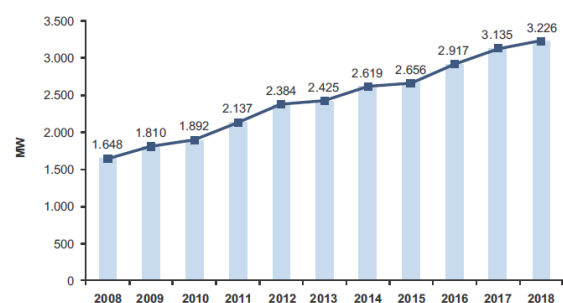


Fig. 1: Histórico de demanda máxima. Fuente: [2].

La distribución de la demanda en el sistema eléctrico paraguayo, mostrada en la Figura 2, indica una fuerte concentración, solo el Sistema Metropolitano (55.4 %). Los pronósticos de la ANDE indican que dicha distribución se mantendría en el corto y mediano plazo, salvo que se incentiven cada vez más complejos industriales de gran porte en localidades del interior del país, que puedan cambiar esta estructura de distribución de la demanda [3].

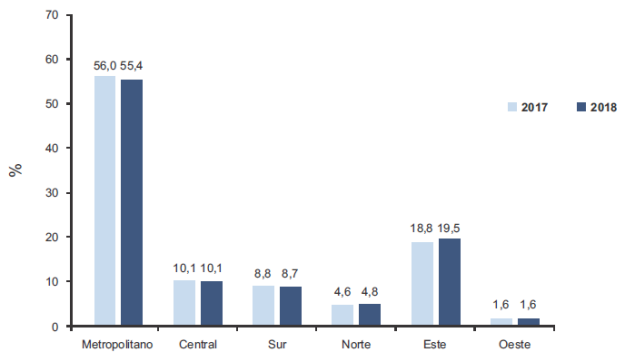


Fig. 2: Histórico de demanda máxima. Fuente: [2].

- **Demanda anual:** esta íntimamente relacionada con el desarrollo de la actividad económica del país y con la producción de bienes y servicios. La serie de la demanda es un indicador de la marcha de la economía y su tendencia a largo plazo. [[12]]
- **Demanda mensual:** Se observa moderadas variaciones entre meses debido a los cambios estacionales a lo largo de año y da lugar a un perfil de demanda mensual típico.
Dados los efectos incidentes de las variables meteorológicas, el consumo en verano y los meses próximos al mismo es altamente superior a la media anual mientras que en invierno suele ser inferior. A medida que la temperatura se acerca a sus máximas usuales a lo largo del año, la disposición a utilizar aparatos de climatización es mayor.
- **Demanda diaria:** La evolución de la demanda eléctrica diaria sigue un ciclo semanal y esta caracterizada por la actividad laboral o productiva propia de cada día de la semana [12].
- **Demanda horaria:** El perfil horario de demanda sufre fuertes variaciones y presenta dos o tres picos durante el día y un vasto valle por la noche y madrugada. Sin embargo, se puede identificar un patrón estable de dicho perfil horario en función al tipo de día y de factores como la laboralidad o el clima.

2.2. Operación del Sistema Interconectado Nacional

El Sistema Interconectado Nacional (SIN) es el sistema eléctrico de potencia destinado al abastecimiento de la demanda de energía eléctrica de consumo interno del país, mientras que el Sistema Interconectado (SI) es el sistema eléctrico de potencia destinado al abastecimiento de la demanda de energía eléctrica de consumo tanto interno como de las interconexiones destinadas a exportaciones.

En resumen, el valor cuantitativo del SIN en un determinado tiempo t , se puede representar de la siguiente forma:

$$SIN(t) = SI(t) - Ventas$$

$$SI(t) = P_{IPU}(t) + P_{YAC}(t) + P_{ACY}(t) + Compras$$
(1)

El SIN como potencia eléctrica (normalmente en el orden de MW), es la principal variable de estudio en el presente Trabajo Final de Grado, que es de suma importancia su predicción a ciertos plazos que son decisivos para la toma de decisiones en las centrales de operaciones de la ANDE.

Su importancia radica, tanto del SI como el SIN, en la búsqueda de la optimización de la continuidad y calidad del servicio integrado de los sistemas eléctricos de generación, transmisión y distribución de energía eléctrica de la ANDE. También se convierte en el principal parámetro de análisis para acuerdos contractuales de intercambio de energía con las dos entidades binacionales [3].

Como consecuencia del crecimiento de la demanda de energía a lo largo de los años, el SIN se encuentra vulnerable a problemas de inestabilidad de tensiones, confiabilidad y disponibilidad del suministro de energía eléctrica. [1].

2.3. Metodología Actual y Planteamiento del Problema

La programación energética diaria realizada por el departamento correspondiente de la ANDE con base en experiencia de los especialistas y con métodos estadísticos tradicionales de previsión, estiman la curva de carga del sistema para cada día [3]. Una vez determinado así el perfil de demanda, la atención se enfoca en la potencia contratada del mes en Itaipú, la disponibilidad de maquinarias en la CH-ACY, la disponibilidad de potencia excedente en la CH-IPU, el salto neto de itaipú y la configuración del SIN.

En resumen, la problemática que el presente Trabajo Final de Grado pretende resolver es la automatización del proceso de programación de carga diaria dirigidas por el despacho de potencia de la ANDE. En términos de largo plazo, servirá para optimizar las decisiones de las compras contractuales de potencia cuando el sistema eléctrico nacional este totalmente conectado, teniendo así un mejor balance energético entre el suministro y demanda eléctrica a nivel nacional.

3. Marco Teórico

3.1. Serie temporal del Consumo Eléctrico

El consumo de energía eléctrica es una serie temporal que comúnmente es de alta aleatoriedad debido a las diversas operaciones eléctricas y del comportamiento del consumidor que varía con el tiempo, lo que hace que la identificación de los patrones de esta serie temporal lo vuelve mucho más difícil y compleja [[6]].

Las series temporales permiten distintas clasificaciones dependiendo del aspecto en el que se centre la atención:

- **Series univariantes:** Solo se dispone de una variable a estudiar en función del número de respuestas que recoge el valor que adopte dicha variable.
- **Series multivariantes:** Presentan más de una variable objeto de estudio y en cada periodo de observación que recoge el valor de cada una de dichas variables.
- **Series estacionarias en media:** En este tipo de series los valores oscilan en torno a una media constante con una variabilidad que también se mantiene constante con respecto a dicha media a lo largo del tiempo. Son también denominados series temporales estables.
- **Series estacionarias en varianzas:** En estas series la mayoría de los valores se encuentran a la misma distancia respecto a la media, es decir, presentan un grado de dispersión constante en el tiempo. También llamado homocedasticidad.

- **Serie estacionales:** Presenta un comportamiento estacional si presenta una tendencia que se repite cada cierto periodo de tiempo.

Una serie temporal univariante, como lo es la demanda eléctrica, se puede descomponer en tres componentes básicas lineales y no lineales, según su variación: Estas componentes son:

- **Tendencia:** Hace referencia a los cambios que se producen en el largo plazo en relación a la media de la serie. Puede ser ascendente, descendente o ausente. Se caracteriza por inducir un movimiento suave de la serie a largo plazo.
- **Estacionalidad:** Esta componente induce un patrón en la serie que se repite cada cierto periodo de tiempo. Un ejemplo de como la estacionalidad influye en una serie es el aumento del consumo de energía eléctrica en los meses de verano debido a las altas temperaturas y en consecuencia, al uso de aparatos de refrigeración.
- **Componente aleatoria o Residuo:** Tras eliminar de la serie la tendencia y la estacionalidad se obtienen unos valores aleatorios.

3.2. Modelos estadísticos

Para la elaboración del modelo predictivo con técnicas estadísticas, se adapta la metodología clásica de Box-Jenkins, donde se identifica la naturaleza de la serie temporal, se realiza una estimación cuantitativa del modelo y como última fase, se realiza la predicción [10].

3.2.1. Identificación

El modelo a diseñar exige que las series temporales univariantes o multivariantes, sean al menos, estacionarias en media (tendencia constante) y estacionarias en varianza (varianza constante). En consecuencia, se realiza ciertas transformaciones a las series temporales, resumida en Tabla

TABLA 1: Transformaciones series temporales.

Serie Temporal	Transformaciones
No estacionario en media	Diferenciación orden n $z_t = y_t - y_{t-n}$
No estacionario en varianza	Box-Cox o Logarítmica

Fuente: Elaboración Propia.

La estacionariedad de una serie temporal es examinada estadísticamente por la prueba de raíz unitaria, donde **Dickey-Fuller** es el mas usado. La hipótesis nula H_0 es la presencia de una raíz unitaria en la componente estocástica de la serie temporal, que ciertamente indica la no-estacionariedad en la misma.

3.2.2. Estimación

Una vez que la serie sea estable en media y varianza se procede a estimar los parámetros de los siguientes procesos para el modelado

- **Procesos Autorregresivo AR** La dependencia de los valores pasados en una serie temporal da lugar a la aparición de una cierta inercia en la misma a lo largo del tiempo. Los modelos más simples para representar esta inercia son los **procesos autorregresivos**, los cuales aplican la idea de regresión simple con el fin de representar la dependencia lineal en un instante de un tiempo t entre una variable aleatoria Y_t y esta misma variable en un momento anterior Y_{t-1} .
- **Procesos Media Móvil MA** El proceso de media móvil se puede escribir como una función de todas las nuevas informaciones generadas por el proceso autorregresivo, tendiendo a cero con el retardo. Sin embargo, los procesos AR no pueden representar series de muy cortas memoria, es decir, series donde el valor actual esta correlacionado con un numero reducido de valores anteriores. Para este caso, en el que el valor actual es función de un numero pequeño de innovaciones pasadas se emplean los procesos de **media móvil** o procesos MA.
- **Procesos Integrados I** Los modelos autorregresivos integrados de media móvil o ARIMA (p,d,q) son una combinación de todos los procesos descriptos anteriormente con la incorporación del Proceso Integrado I(d), donde d representa el numero de diferencias necesarias para obtener un proceso estacionario a partir de un proceso no estacionario en media, encontrado en la mayoría de las series temporales.
- **Modelo SARIMAX** Modelo Estacional Autorregresivo de Medias Móviles Integrado con variables exógenas. Es una extensión lógica del modelo ARIMA, que incorpora variables independientes de naturaleza explicativa y además tiene en cuenta la componente estacional de la serie temporal. Conceptualmente, es una fusión de los modelos de regresión y del modelo ARIMA. Cuando los termino AR y MA de un modelo ARIMA puro no son suficientes para obtener pronósticos aceptables, es conveniente investigar algún otro fenómeno, cuya influencia en el tiempo no esta suficientemente relacionada en los datos históricos de la serie temporal. A la hora de modelar series con dependencias estacional y regular esencialmente se considera ambas separada y de forma multiplicativa. Este modelo es identificado por sus componentes (P,D,Q) y por el periodo de tiempo s que marca el numero de observaciones que componen el ciclo estacional.

3.3. Modelos de aprendizaje de máquina

La predicción o el *forecasting* es uno de los abordajes más importantes de los modelos de aprendizaje de maquina o *Machine Learning*. Actualmente se desarrollan algoritmos para adecuarse a la complejidad de proyección en el futuro. Es por eso que las regresiones del *Machine Learning* nos ofrecen mayores ventajas con respectos a técnicas más tradicionales: velocidad en el preproceso de datos, identificación de patrones ocultos en los mismos, mayor exactitud en la predicción y la creación de un sistema más robusto que es capaz de actualizarse y adaptarse ante nuevos datos o *feataures* [6].

Los regresores de *Machine Learning* mas aplicados en la literatura son: Regresion Lineal, *Random Forest*, *Extra Trees*, *K-nearest neighbors*, *Gradient Boosting*, entre otros.

3.4. Modelos de aprendizaje profundo

El aprendizaje profundo, o bien llamado *Deep Learning* no es más que un conjunto de algoritmos de aprendizaje automático que intenta modelar abstracciones de alto nivel en datos usando arquitecturas computacionales que admiten transformaciones no lineales múltiples e iterativas de datos expresados en forma matricial o tensorial. Estas abstracciones en el modelo son posibles gracias a las redes neuronales artificiales [9].

En particular, las *Redes Neuronales Recurrentes* son especializadas en procesar datos secuenciales, $x(1), \dots, x(t)$ de longitud variable y con la capacidad de incrementar el tamaño de las secuencias, algo inviable para redes sin especialización basada en secuencias.

3.4.1. LSTM

Long-Short Term Memory, conocido como **LSTM** forma parte del extenso campo de las redes neuronales recurrentes, situados dentro de los modelos de *Deep Learning*. Esta constituido por un bloque de memoria llamado celda LSTM. En la Figura 3 se observa como es una celda LSTM por dentro:

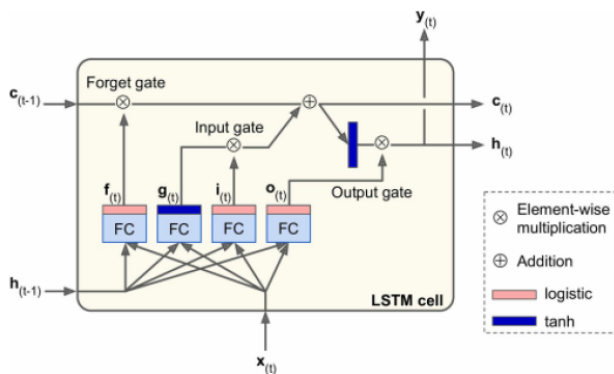


Fig. 3: Célula LSTM. Fuente: [9]

En resumen, una celda LSTM puede aprender a reconocer datos *inputs* importantes (cuyo rol es de la puerta de entrada), almacenarla en el estado de largo plazo, aprender a guardarla siempre que sea necesaria (cuyo rol es el de la puerta de olvido) y aprender a extraer su información siempre y cuando necesite. Todo esto es la razón de por que LSTM ha tenido tanto éxito a la hora de encontrar patrones a largo plazo en series temporales, largos textos, grabaciones de audio, etc. Además de corregir el problema del desvanecimiento del gradiente, que es común dentro de las redes neuronales recurrentes [7].

3.5. Optimización de hiperparámetros

3.5.1. Algoritmo Genético

La forma mas segura de prevenir el *overfitting* o el *underfitting* durante el entrenamiento del modelo, es la de encontrar los mejores valores de los hiperparámetros de la red. Modelos como [8] o [6] utilizan el algoritmo genético con las mas convenientes configuraciones, pero requieren grandes cantidades de procesamiento de datos y de tiempo computacional. [4].

4. Metodología aplicada

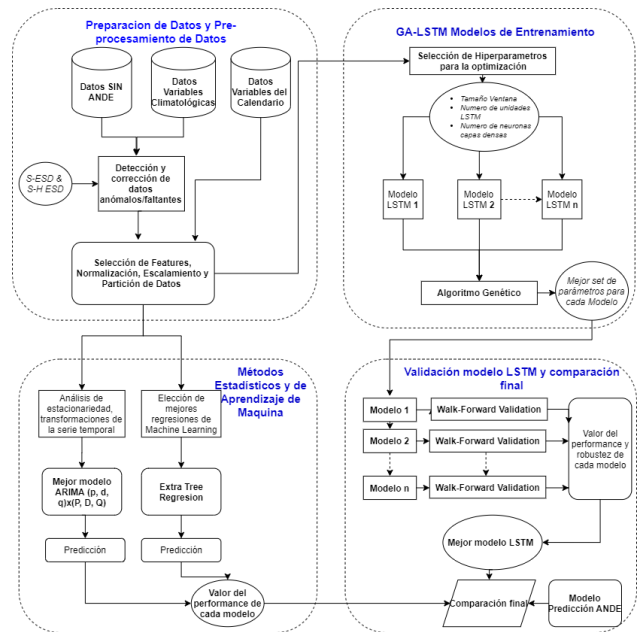


Fig. 4: Overview de la Metodología Aplicada. Fuente: Elaboración Propia

4.1. Preparación de datos y pre-procesamiento

4.1.1. Preparación de datos

Todos los datos empleados en este Trabajo Final de Grado, son datos oficiales de entidades públicas. El conjunto entero de datos que corresponde al consumo eléctrico nacional, el cual posee un registro horario, desde el año 2008 hasta el año 2018.

Unos de los *features* más incidentes con el consumo eléctrico son las variables meteorológicas. Los datos reales usados en el modelo, proceden de entidades públicas diferentes: El Centro Meteorológico de la Facultad Politécnica de la Universidad Nacional de Asunción y la Dirección de Meteorología e Hidrología.

Las variables meteorológicas disponibles fueron la temperatura seca del aire, la humedad relativa, velocidad del viento y la dirección del viento de once estaciones meteorológicas distribuidas en diferentes ciudades de diferentes departamentos. A partir de estos datos se ha unificado todo en una sola variable, *Temperatura Aparente*, obtenida mediante la fórmula de Steadman [14].

También se ha considerado el efecto de las variables del calendario sobre el consumo eléctrico, como el año, meses, trimestres, día de la semana y horas.

4.1.2. Pre-procesamiento

Limpieza de datos Se han encontrado diversas irregularidades en las serie temporales. El mayor inconveniente fue la falta de datos en cortos y grandes intervalos, debido a la fina escala de tiempo seleccionada. El segundo problema que se tuvo fue la existencia de datos anómalos o atípicos (*outliers*) en la serie temporal del SIN y de las variables meteorológicas.

La detección y corrección de datos anómalos, ha estado a cargo de los métodos de *Seasonal Extreme Studentized Deviate S-ESD* y el *Seasonal-Hybrid Extreme Studentized Deviate S-H-ESD*, propuesto por [11]. Se ha tomado la variante del residuo original, debido a que la serie temporal presenta una pronunciada tendencia creciente.

Finalmente, se ha seleccionado el *S-H-ESD* tanto para la detección de *outliers* de la serie temporal del SIN y de las variables meteorológicas, a pesar de tener un mayor número de falsos positivos, debido a que se han encontrado una gran cantidad de anomalías pertenecientes a errores locales los cuales no pueden ser percibidos por el primer método, luego de una exhaustiva revisión.

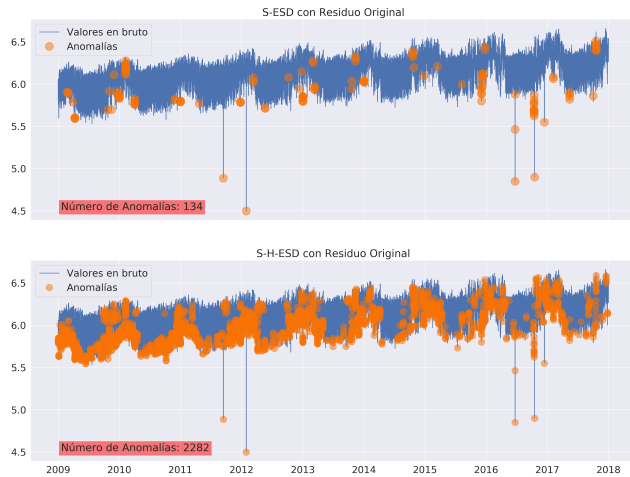


Fig. 5: Detección de *outliers* en el SIN. **Fuente:** Elaboración Propia.

Una vez detectado y eliminado los datos anómalos en la componente residual, se ha procedido a sumar esta componente con la componente de la tendencia, cuyo resultado se conoce como ajuste estacional, sobre la cual se aplicó una interpolación lineal para rellenar los datos faltantes.

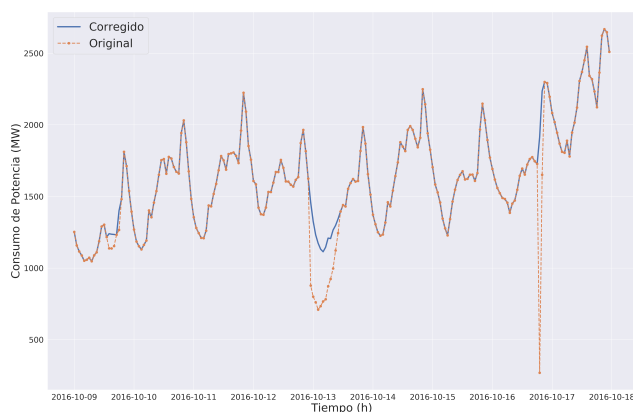


Fig. 6: Corrección de *outliers* en el SIN. **Fuente:** Elaboración Propia.

Selección de features mediante la correlación *Pearson* entre el SIN y todas las variables numéricas relevantes posibles, plasmado en una mapa de calor, como en la Figura 7. Además también en la Figura 8 donde muestra las mejores correlaciones del SIN de horas anteriores con el SIN de una hora actual.

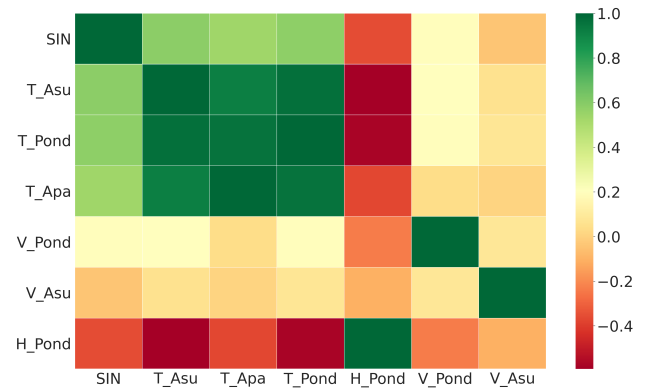


Fig. 7: Mapa de correlación. **Fuente:** Elaboración Propia.

La Figura 7 enseña que la temperatura de Asunción es la que tiene mayor correlación con el SIN, dejando atrás la temperatura aparente. Esto se debe por dos razones: en Asunción esta concentrada la mayor parte de la carga del sistema eléctrico paraguayo (Figura 2) y cuya serie temporal fue la que presentó menor cantidad de datos faltantes y anómalos. En consecuencia, solo se ha utilizado la temperatura de Asunción como única variable exógena para el *input* del modelo.

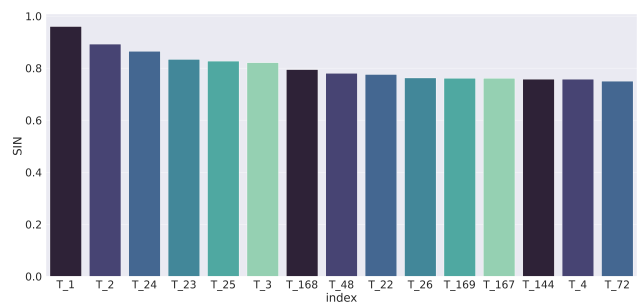


Fig. 8: Histograma de correlaciones del SIN. **Fuente:** Elaboración Propia.

4.2. Modelos ARIMA

Primeramente se debe estudiar la estacionariedad de la serie temporal, antes de estimar los parámetros p y q , para los procesos autorregresivos y media móvil, respectivamente.

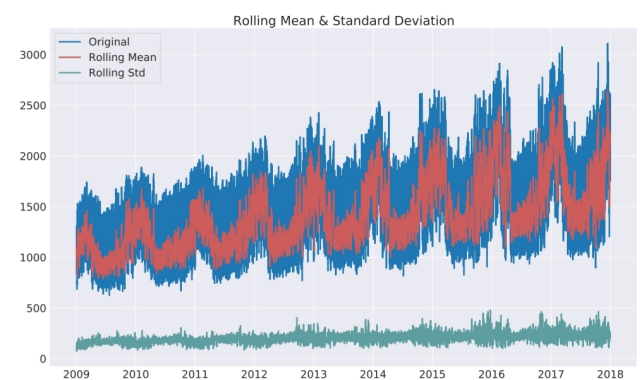


Fig. 9: No-estacionariedad en media de la serie temporal. **Fuente:** Elaboración Propia.

Seguidamente se dispone a realizar la diferenciación de orden uno para obtener la estabilidad en media en la serie temporal, como enseña la Tabla 1. Luego, se estima el mejor conjunto de parámetros apoyándose en los correlogramas de

las funciones *ACF* y *PACF*, mostrados en la Figura 10. En consecuencia, se elige el modelo *ARIMA*(2, 1, 3) como mejor modelo generador de la serie temporal. Además se incluye el efecto estacional en la serie temporal y a la Temperatura de Asunción como variable exógena, obteniéndose finalmente el modelo *SARIMAX*(3, 1, 2)(1, 0, 1).

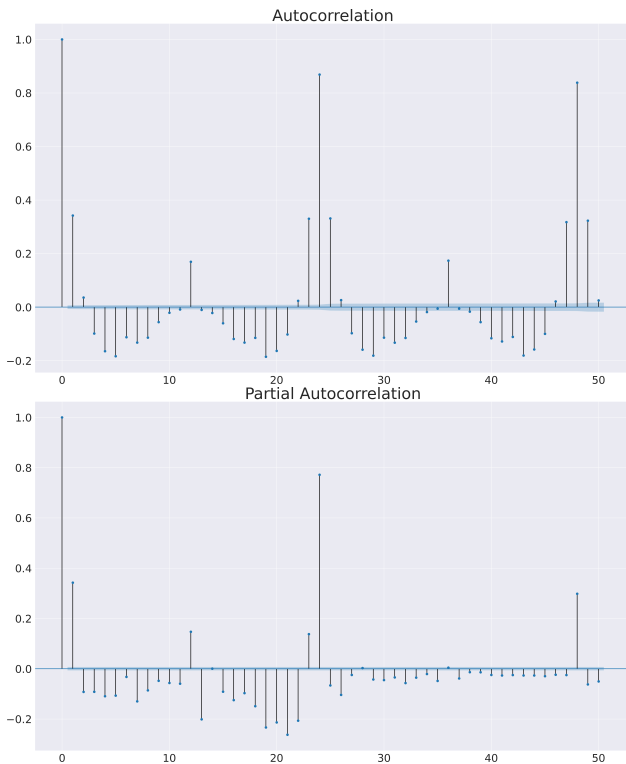


Fig. 10: Funciones ACF y PACF. Fuente: Elaboración Propia.

4.3. Modelos de Machine Learning

Se ha seleccionado a *Random Forest* como mejor regresor en la tarea de predicción el consumo eléctrico. Los parámetros del mismo fueron optimizados mediante una *Grid Search*, cuyos resultados fueron:

TABLA 2: Parámetros del regresor *Random Forest*.

Parámetros	Valor
Número de arboles	75
Máxima profundidad de árbol	100
Máximo Numero <i>Features</i>	$\log_2$

Fuente: Elaboración propia.

4.4. Modelo LSTM

El modelo LSTM propuesto posee una capa de salida de 24 neuronas correspondientes a las 24 horas del día que se desea predecir, comenzando siempre de la hora 0 hasta las 23 horas. Esto permite un mayor enfoque sobre la curva de demanda por día, evitando generalizaciones que podrían conllevar a errores más notorios.

Con respecto a los *features* a ser utilizados en este modelo, se ha tenido en cuenta las Figuras 7 y 8 para la determinación de los mismos.

Teniendo en cuenta las variables del calendario, se tomaron en consideración el año, el trimestre, el mes, el día del mes, el día de la semana y la hora del día.

TABLA 3: Resumen de *Features*.

<i>Features</i>	Descripción
SIN Actual	Valor en t
Temperatura de Asunción Actual	Valor en t
Año Actual	Valor en t
Mes Actual	Valor en t
Día de la semana Actual	Valor en t
SIN una semana Anterior	Valor en $t-168$
Temperatura de Asunción Posterior	Valor en $t+24$
Año Posterior	Valor en $t+24$
Mes Posterior	Valor en $t+24$
Día de la semana Posterior	Valor en $t+24$

Fuente: Elaboración propia.

Posterior a la selección de los *features* más relevantes para el modelo, se llevó a cabo la elección de los hiperparámetros de la red LSTM, resumidas en la Tabla 4

TABLA 4: Hiperparámetros de una red LSTM.

Hiperparámetros	Candidatos
Tamaño de ventana	Entre 1 y 168
Número de capas LSTM	Una
Número de unidades recurrentes LSTM	1 hasta 100
Número de capas densas	Entre 1 y 6
neuronas en capas densas	Entre 10 a 200
Optimizador	ADAM
Función de activación	ReLU
Función de costo	MSE

Fuente: Elaboración Propia.

Los hiperparámetros que se han decidido optimizar debido al alto impacto en el desempeño de la red son:

- **Tamaño de ventana** Es el mas importante para un modelo LSTM ya que el modelo será capaz de capturar las dependencias condicionales entre periodos de tiempo sucesivos [5].
- **Número de neuronas en las capas densas** El algoritmo metaheurístico será capaz de encontrar el número óptimo en cada capa densa, especialmente en las últimas ya que aportan una gran influencia en la capa de salida.
- **Número de unidades recurrentes LSTM** Se ha elegido debido a la dificultad de elegir un número exacto dada la cantidad de capas densas. Estos dos últimos hiperparámetros son importantes para la robustez de la red [6].

Dado que se ha seleccionado una sola capa LSTM, se decidió variar el número de capas densas, construyendo 6 modelos diferentes con configuraciones de 1 hasta 6 capas densas ocultas.

TABLA 5: Resultados finales del Algoritmo Genético.

Ventana	Celdas LSTM	Número neuronas	RMSE
101	53	179	0.04052
117	92	106, 44	0.03898
117	92	152, 55, 93	0.04010
118	88	82, 77, 186, 126	0.03815
107	94	152, 15, 23, 59, 192	0.03968
101	53	179, 39, 36, 74, 22, 144	0.03904

Fuente: Elaboración propia.

Para estudiar la robustez de los modelos se ha llevado a cabo el proceso de validación bajo los conceptos del *Walk-Forward* con las dos variaciones: Ventana Expandida (*Expanding Window*) y Ventana deslizante (*Sliding Window*) con diez iteraciones cada una.

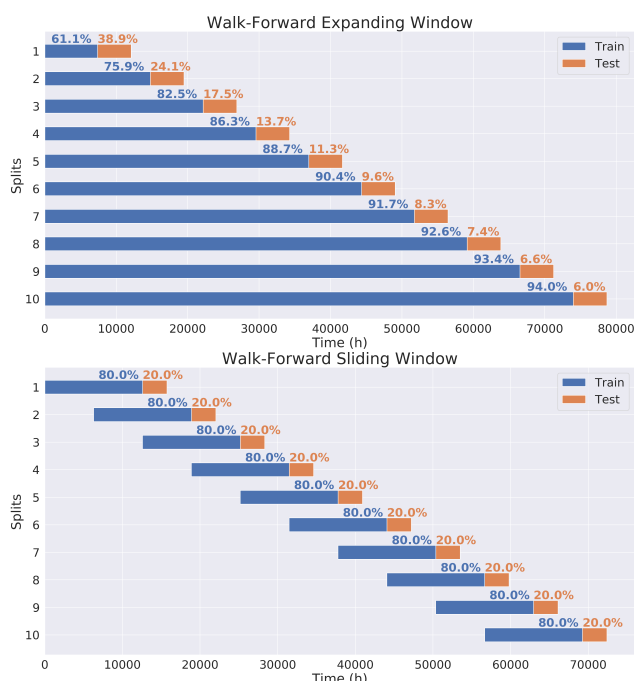


Fig. 11: Particiones del *Expanding Window* y *Sliding Window*. Fuente: Elaboración Propia.

Para ambos tipos de validaciones se comprobó la estabilidad de los errores de los seis modelos durante las diez iteraciones para asegurarse la robustez y generalización de las mismas, mostrando independencia de las particiones entre datos de entrenamiento y prueba.

5. Resultados y Comparaciones Finales

Para un mejor panorama de los análisis finales, se ha realizado la desnormalización de los valores del conjunto de prueba abarcando el ultimo año del conjunto entero de datos. En la Tabla 6 muestra que el modelo 2 es el se destaca en todas las métricas.

TABLA 6: Métricas de los modelos para todo el conjunto de prueba

Modelo	RMSE	MAE	MAPE
1	103.45	72.68	4.16
2	91.65	66.71	3.8
3	94.5	69.42	4.02
4	96.34	71.75	4.17
5	99.65	73.21	4.23
6	98.02	68.76	3.98

Fuente: Elaboración propia.

TABLA 7: Evaluación del modelo 2 (LSTM) frente al modelo estadístico y de aprendizaje automático

Modelo	RMSE	MAE	MAPE
LSTM	91.65	66.71	3.8
Ramdon Forest	249.91	191.24	10.65
SARIMAX	422.61	357.82	22.78

Fuente: Elaboración propia.

Mediante la Tabla 7 se puede concluir que el modelo LSTM obtiene mejores resultados frente a los demás métodos propuestos, en el mismo conjunto de prueba.

Un análisis más general del modelo puede visualizarse en los histogramas de la Figuras 12, en donde enseña que el mayor consumo se produce con más frecuencia a las 18 horas, seguido de las 21, 20 y 19 horas, en ese orden; con una variación del error mayormente entre los 0 y 100 MW.

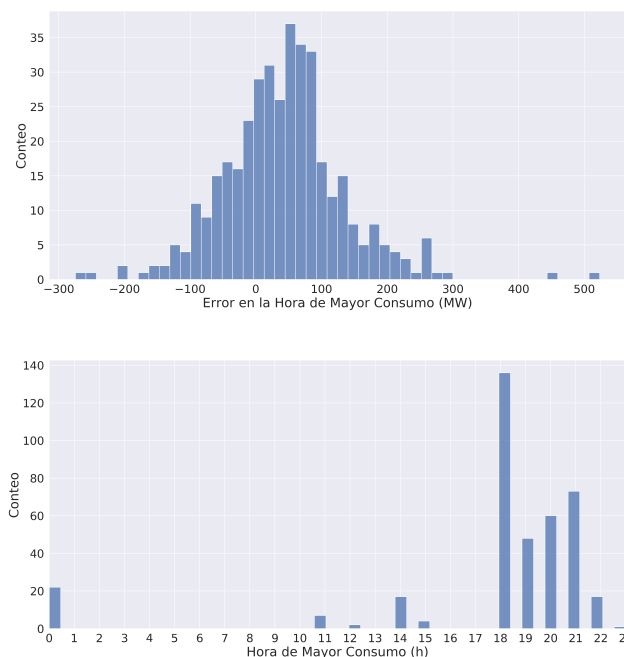


Fig. 12: Histogramas del Error en la Hora de Mayor Consumo y la Hora de Mayor Consumo. Fuente: Elaboración propia.

Finalmente se realiza una comparativa final con las predicciones realizadas por ANDE. Se eligió el día 05-11-2015,

ya que fue uno de los días en la que ANDE obtuvo mayor precisión.

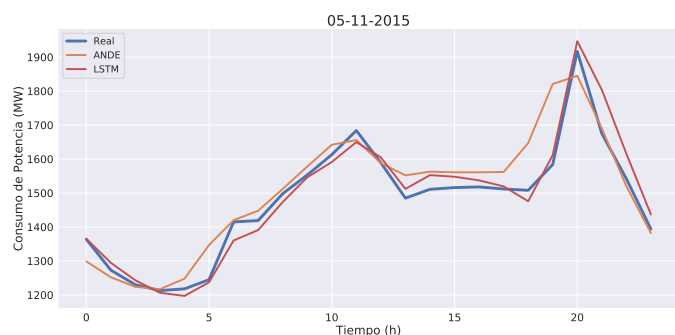


Fig. 13: Previsión del modelo LSTM frente a la ANDE.
Fuente: Elaboración Propia.

TABLA 8: Métricas de la previsión del modelo LSTM frente a la ANDE del día 05-11-2015

Modelo	RMSE	MAE	MAPE
LSTM	39.79	30.01	1.98
ANDE	69.05	46.37	3.11

Fuente: Elaboración propia.

La Figura 13 y la Tabla 8 dejan evidencias que las predicciones del modelo LSTM propuesto obtiene mejores resultados comparados con las programaciones en la ANDE.

Por este motivo se propone la implementación del modelo desarrollado, en el despacho de carga de la ANDE, con el fin de proporcionar un apoyo en la toma de decisiones que aparecen en el momento de la provisión de demanda para corto plazo (24 horas del día siguiente). Esto permitiría tener una opción más amplia en la estrategia de actuación, ya que junto con las técnicas tradicionales utilizadas según la situación lo requiera, el operario podrá visualizar las distintas situaciones y estimaciones que podrían darse.

Los mecanismos necesarios a tener en cuenta para la implementación del modelo en el despacho de carga de la ANDE, consisten en los siguientes pasos:

1. Recopilación y estructuración del histórico de demanda eléctrica y de la temperatura de Asunción de fuentes oficiales, como así también del pronóstico de temperatura de los días a ser predichos, de acuerdo a las exigencias requeridas por el modelo.
2. Creación de una base de datos para el almacenamiento y actualización de los mismos, de manera a facilitar el rápido acceso a las variables precisadas por el modelo.
3. Creación de una interfaz capaz de acceder a la base de datos y realizar las predicciones solicitadas, además de la visualización de otros parámetros que el operador del despacho de carga considere relevantes.

6. Conclusiones y Recomendaciones

6.1. Conclusiones

Las estrategias utilizadas para la recopilación y preprocesamiento de los datos históricos de la demanda eléctrica y de las variables correlacionadas a éste, han demostrado cumplir efectivamente con la identificación de

anomalías y relleno de datos faltantes, posibilitando un correcto proceso de aprendizaje de los modelos desarrollados.

Los modelos SARIMAX y Random Forest, correspondientes a los modelos estadísticos y de aprendizaje automático respectivamente, se desempeñaron de manera correcta demostrando la capacidad para realizar predicciones que sirven de base para analizar otros métodos más complejos enfocados en aprendizaje profundo.

La generación de varios modelos basados en LSTM, con distintas configuraciones y cuyos parámetros fueron optimizados de manera satisfactoria mediante la utilización de un algoritmo genético diseñado específicamente para esta tarea, ha permitido obtener los mejores candidatos para un análisis posterior.

Con el proceso de validación y los análisis realizados bajo criterios y exigencias de un sistema de predicción de demanda eléctrica a nivel país, el modelo LSTM se ha destacado en precisión y robustez.

La comparativa final realizada con el modelo LSTM elaborado en este Trabajo Final de Grado demuestra mejores resultados en la predicción de demanda eléctrica frente a los modelos SARIMAX y Random Forest, como así también, frente a las predicciones realizadas por la ANDE para los días analizados, en donde sobresale la precisión en las horas de mayor consumo.

6.2. Recomendaciones

El presente Trabajo Final de Grado ofrece una base adecuada para un sistema integral de predicción de consumo eléctrico, con el fin de ser replicada para diseños de modelos más reproducibles en cuanto a particularidades propias de cada día de la semana o trimestre. También modelos que puedan predecir comportamientos atípicos en el día o en la semana.

En cuanto al horizonte de la predicción del modelo, puede ser cimiento para la elaboración de otros con proyecciones de mediano y largo plazo, con tal de la inclusión de variables adecuadas que acompañen los patrones observados en el consumo eléctrico. Los mismos servirían como base para la toma de decisiones para la optimización de contrato de potencia de la Itiipu Binacional.

7. Referencias

- [1] ANDE (2016). Plan maestro de generación.
- [2] ANDE (2018). Memoria anual 2018. *Nombre de la revista*.
- [3] Barboza, O. (2014). Automatización de despacho horario de potencia eléctrica en el sistema interconectado nacional del Paraguay.
- [4] Bodenhofer, U. (2003). Genetic algorithms: theory and applications.
- [5] Bouktif, S., Fiaz, A., Ouni, A., y Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, **11**(7): 1636.
- [6] Bouktif, S., Fiaz, A., Ouni, A., y Serhani, M. A. (2020). Multi-sequence lstm-rnn deep learning and metaheuristics for electric load forecasting. *Energies*, **13**(2): 391.
- [7] Cabezon, M. (2018). Predicción demanda eléctrica española. implementación de redes neuronales recurrentes en python.
- [8] Chung, H. y Shin, K.-s. (2018). Genetic algorithm-optimized long short-term memory network for stock market prediction. *Sustainability*, **10**(10): 3765.
- [9] Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.
- [10] González Casimiro, M. P. (2009). Análisis de series temporales: Modelos arima.
- [11] Hochenbaum, J., Vallis, O. S., y Kejariwal, A. (2017). Automatic anomaly detection in the cloud via statistical learning. *arXiv preprint arXiv:1704.07706*.

- [12] Jiménez, D. F., Ruíz, J. J., y Huertas, E. C. (2016). Modelo de predicción de la demanda eléctrica horaria a muy corto plazo: Aplicación al sistema peninsular español.
- [13] Kuster, C., Rezgui, Y., y Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. *Sustainable cities and society*, **35**: 257–270.
- [14] Steadman, R. G. (1984). A universal scale of apparent temperature. *Journal of Climate and Applied Meteorology*, **23**(12): 1674–1687.