

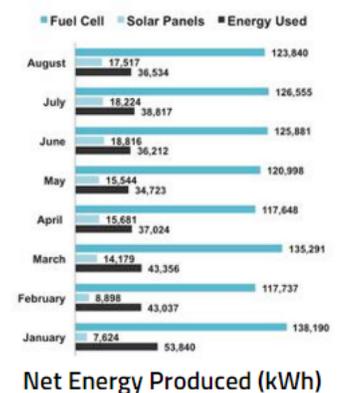
Estancia en Ames Research Center.

Mario Manuel Gurrola Morga.

Llegué al centro de investigación Ames el viernes 02 de octubre y contacté a la líder de la oficina de estudiantes internacionales, **Desireemoi Bridges**, para agendar una cita con mi mentor, el **Dr. Rodney Martin**. Se llegó al acuerdo que mi mentor y yo tendríamos una charla el lunes 05 de octubre de 1 PM a 2 PM. En dicha charla el Dr. Martin me dio una breve introducción a la NASA, a los diferentes campus y en específico al centro de investigación Ames y cuál era su enfoque. Además, me dio un tour por el edificio sustentable y explico el propósito y funcionamiento del mismo.

Sustainability Base (Base de sustentabilidad en inglés) es un edificio verde con certificación *LEED Platino*, esta certificación cuyo nombre en inglés es *Leadership in Energy and Environmental Design* (Líder en Energía y Diseño Ambiental) reconoce y califica el avance y rendimiento de un edificio ecológico en cuatro niveles diferentes (bronce, plata, oro y platino de manera incremental) tomando en cuenta cinco características:

- Sitios sustentables.
- Eficiencia en el uso del agua.
- Uso de energía y atmósfera.
- Materiales y recursos.
- Calidad ambiental dentro del edificio.



Sustainability Base, al tener una certificación de platino resulta ser un gran avance en cuanto a edificios ecológicos. Cuenta con un terreno de 50,000 pies cuadrados de superficie en los cuales se encuentra el edificio en sí con todos sus sistemas internos, pilas de combustible y celdas solares. El propósito de la base sustentable no es solo ahorrar recursos, sino además generarlos. En la gráfica de la derecha se muestran los resultados energéticos publicados por NASA, se puede notar ampliamente que se genera mucha más energía de la que se gasta.

Los sistemas internos del edificio incluyen sistemas de reciclaje de agua, sistemas de ventilación natural, sistemas de calentamiento por medio de bombas de agua caliente, entre otros.

Las dos semanas siguientes a la introducción por parte del Dr. Martin fueron de entrenamientos de seguridad para los laboratorios. Estos entrenamientos cubrían todas las áreas y laboratorios, y hacían énfasis en la seguridad informática.

La tercera semana después de mi llegada (la semana del 19 de octubre) comencé a inmiscuirme en el trabajo que las demás personas hacían respecto al proyecto mediante una serie de correos en los que se me incluía con los resultados y cambios realizados. Sin embargo, yo todavía necesitaba comprender la herramienta a grandes rasgos para entender los problemas y objetivos, y comenzar a trabajar en conjunto.

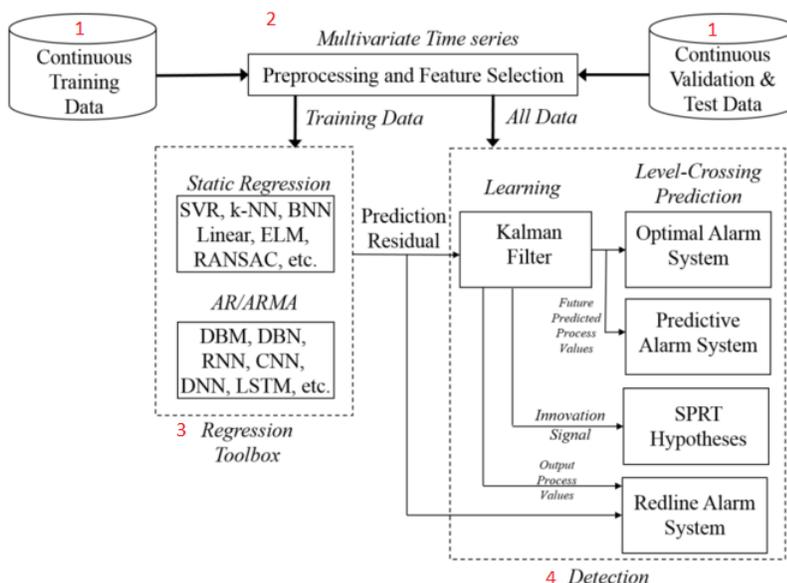
El 31 de octubre, con una comprensión intermedia del trabajo a realizar, pero sobre todo de la herramienta que se utiliza, comencé a revisar el trabajo de los internos anteriores bajo las indicaciones del Dr. Martin.

El 02 de noviembre acudí como invitado a la semi final de *Fame Lab*, un concurso estilo *American Idol* de científicos jóvenes que exponen sus investigaciones de manera que el público en general entienda su trabajo.

Cabe mencionar que el trabajo de la oficina de educación era proveernos a los internos conocimiento propio de la investigación espacial, por lo cual, el jueves 05 de noviembre consiguieron que el **Dr. Seth Shostak**, director del instituto de investigación *SETI* (Búsqueda de Inteligencia Extraterrestre) diera una plática de su trabajo y nos compartiera su discernimiento acerca de la vida inteligente en el universo.

El 06 de noviembre obtuve acceso a los servidores de la NASA con lo cual pude comenzar a trabajar en el proyecto utilizando *ACCEPT*, acrónimo para **Adverse Condition and Critical Event Prediction Toolbox** (Caja de Herramientas para la Predicción de Condiciones Adversas y Eventos Críticos). *ACCEPT*, como su nombre indica, ayuda a predecir eventos anómalos para tomar medidas preventivas.

ACCEPT contiene un amplio repertorio de métodos de aprendizaje de computadora no supervisados, de regresión y de clasificación. *ACCEPT* está desarrollado en **Matlab** y utiliza información pertinente de los 739 sensores físicos dentro de *Sustainability Base* que es recopilada con **Python**.



La arquitectura del software, mostrada en la imagen de la izquierda.

Consta de cuatro partes, la primera es la creación de hojas de datos utilizando información real obtenida del edificio, los datos se clasifican en tres secciones:

- Entrenamiento.
- Validación.
- Pruebas.

La segunda parte es la que se indica bajo el nombre de *Multivariate Time Series* (Series de Tiempo Multivariable) el preprocesamiento y la selección de características, en la cual ACCEPT genera un modelo matemático de información normal utilizando las hojas de datos de entrenamiento usando las características especificadas. Las características son parámetros numéricos como la resolución del modelo y el tamaño de las matrices que se utilizarán para la regresión. Mientras más alto sea el valor de cada característica, tomará más tiempo realizar el procesamiento.

La tercera parte es la que se indica en el apartado de *Regression Toolbox* y en ella se realiza validación cruzada (cross validation) para representar una función en base al error cuadrático medio normalizado (normalized means squared error). Además, esta parte de ACCEPT se encarga de transformar la información de forma que se le puedan aplicar pruebas hipotéticas con las hojas de datos de validación.

La diferencia entre la función base y la función de predicción generada en el paso de regresión es utilizada como parámetro en el filtro de Kalman, el cual es la base para el siguiente paso, el sistema de alarmas.

La última parte, que se indica como *Detection Toolbox*, genera un sistema de alarmas que detecta eventos anómalos respecto al modelo base, este sistema se corre con las hojas de datos de prueba y despliega resultados comparativos entre diferentes métodos de regresión y/o construcción de alarmas. El proyecto donde me vi involucrado era llamado *Cold Complaint Scenario* (Escenario de quejas de frío) en el cual utilizando diferentes sensores de *Sustainability Base* (50 sensores de prueba y 124 en total) se intentaba predecir cambios drásticos en la temperatura (eventos anómalos) para regular la temperatura y así evitar quejas por parte del personal (de ahí el nombre del proyecto).

Mi trabajo consistía en la mejora de hiperparámetros y análisis de la información cuando se realizaba un cambio en las restricciones. Los hiperparámetros son valores numéricos utilizados para modificar el rendimiento del modelo predictivo.

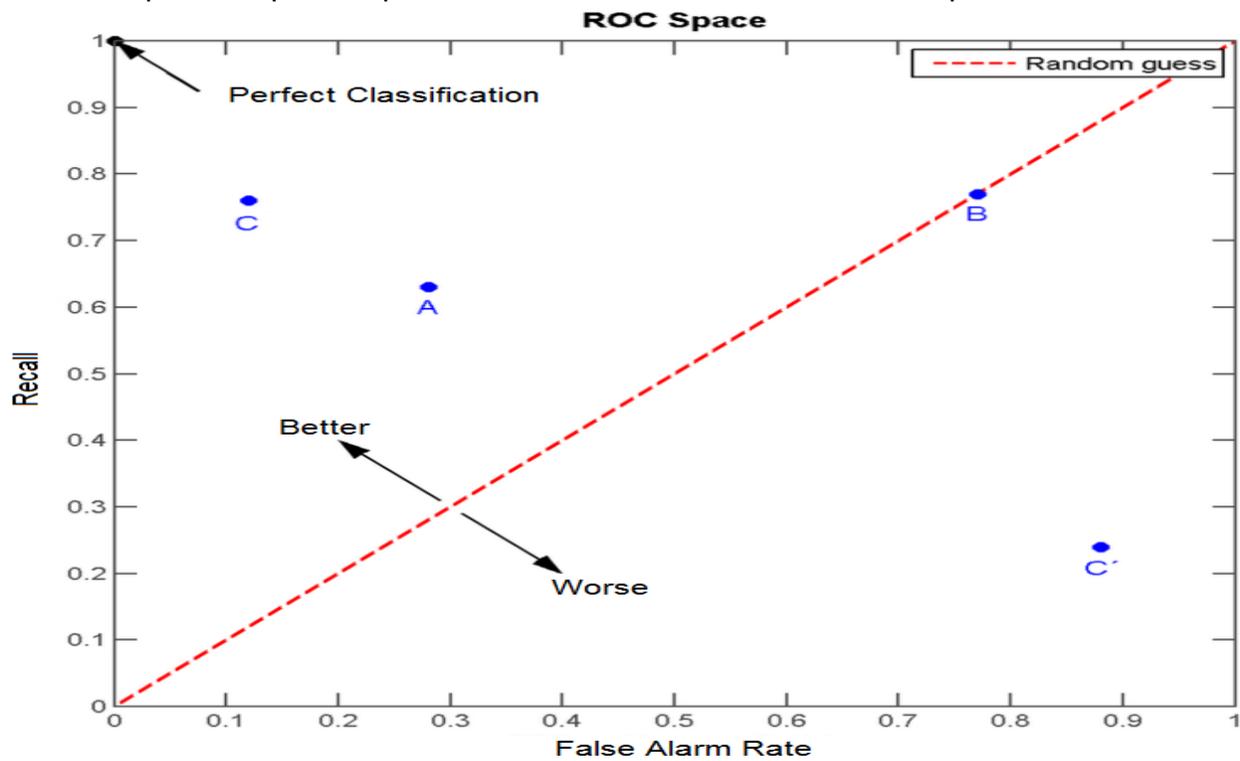
Las restricciones son exactamente lo que su nombre indica, restricciones sobre un parámetro objetivo, dichos parámetros son:

- Missed Detection Rate* (Tasa de detecciones perdidas)
- False Alarm Rate* (Tasa de alarmas falsas)

Para comprender las restricciones es necesario comprender su base, las curvas **ROC** (*Receiver Operating Characteristic*), conocidas como curvas de características operativas del receptor en español, y las matrices de confusión.

Las *matrices de confusión* son clasificadores binarios en los cuales se evalúa si un evento sucedió o no, y si el sistema de alarma dijo que iba a suceder o no, creando así una matriz de dos por dos. Los parámetros objetivo se obtienen como resultado de la clasificación de la matriz de confusión.

Estas curvas son generadas por múltiples matrices de confusión para observar la sensibilidad ante las alarmas falsas y las detecciones perdidas. Se puede restringir una o la otra para obtener un resultado que detecte más, pero de más alarmas falsas o vice versa. También se puede equilibrar para mantener un balance entre ambos parámetros.



Clasificación Perfecta

A

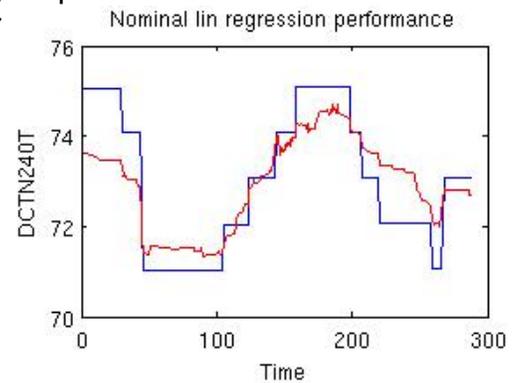
B

C

La coordenada $(1,0)$ es considerada como la clasificación perfecta, pues indica 0% de alarmas falsas y es 100% sensible, es decir, detecta todos los eventos. La línea roja que cruza en medio es la zona del azar, es como lanzar una moneda al aire. En este ejemplo se observa que a pesar de que debajo de la línea del azar los resultados son peores, simplemente se pueden invertir, como se muestra en el caso de C' .

Debido a la amplia variedad de métodos de regresión y clasificación, mi investigación se concentró en los dos métodos más ligeros (menos tiempo de procesamiento y peores resultados) **ELM** (*Extreme Learning Machine*) y **regresión lineal**, aunque el objetivo es verificar esos resultados con métodos más pesados que tienen mayor rendimiento, como la **regresión cuadrática**. Los hiperparámetros de regresión lineal y ELM son coeficiente de regularización y neuronas escondidas (Hidden Neurons) respectivamente.

El objetivo de modificar los hiperparámetros era obtener una función predictiva que describiera el comportamiento de la función original, pero que no fuese igual, pues la función original siempre es entrenada con eventos formales, y la predictiva incluye eventos anómalos.



El viernes 20 de noviembre la oficina de educación invito al astronauta Steven Smith a darnos una plática acerca del trabajo realizado en la estación espacial.

Todo el trabajo que hice en el proyecto de *Cold Complaints Scenario* acerca de restricciones e hiperparámetros fue entregado y presentado a NASA en un póster el 10 de diciembre.



Acknowledgements: Special thanks to the Mexican Space Agency and to my home university. Thanks to Grace Lin for her support, to Rodney Martin and David Nielsen for their guidance and Stefan Hosein for his past work.

Mario Manuel Gurrola Morgia
Zapopan's Superior Institute of Technology
mariogurrolam@gmail.com

Introduction

- Sustainability Base is a LEED-platinum certified high-performance green building.
- The BAS (building automation system) is outfitted with 739 physical sensors which monitor many of the building's environmental conditions and mechanical equipment.
- Data archived from the BAS can be used to provide highly accurate advance warnings of known adverse events.
- The data used for this research is data related to the cold complaint scenario which involves temperature inside target building N232.

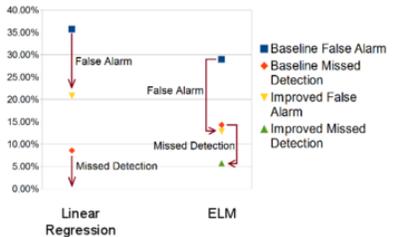


ACCEPT Toolbox

- The toolbox used for adverse event prediction is called ACCEPT (Adverse Condition and Critical Event Prediction Toolbox).
- The purpose of ACCEPT is to compare the performance of a variety of combinations of algorithms.
- ACCEPT builds nominal regression models with training data sets. Deviations from the model can be mapped to forecast anomalous events.
- ACCEPT has a wide array of regression and detection methods that vary in efficiency and resource consumption.

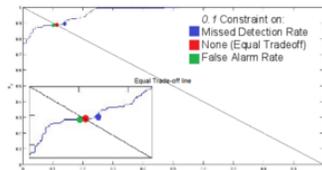
Hyperparameter Improvement

- Hyperparameters are used to tune the performance of the regression model. They heavily affect the results presented by ACCEPT because of their involvement in the creation of the regression model.



Thresholds

- Thresholds can be modified to control the trade-off among the performance metrics: *missed detection rate, false alarm rate and detection time*.
- Thresholds can be selected in the most useful way (prioritizing missed detection, false alarms or balancing them) to improve the output.
- The threshold can be selected using the ROC curve as a guide.



www.nasa.gov

Objective

- Due to cold temperatures inside of the target building N232 in fall and winter, occupant discomfort became a problem.
- The objective of the *cold complaints scenario* is to predict drops in temperature and isolate which parameters may be influencing those drops.
- Using *Stefan Hosein's* work about *cold complaint scenario prediction* and *Benjamin Metzger's* work about *Bayesian networks as baseline* for the research, it is possible to improve the performance of ACCEPT.
- Hyperparameter improvement can influence these forecasts by creating an appropriate statistical fit between various subsets of input parameters to the changes of temperature over a period of time.
- Threshold improvement can change the perspective of the results to favor one performance metric over another.

Conclusion

- Comparing the improved hyperparameters that yielded the best results against the regular hyperparameters the following results were obtained:

Type	False Alarm	Missed Detection	Detection Time
Baseline Linear Regression	35.77%	8.57%	10.5 s
Baseline ELM	28.98%	14.29%	1s
Improved Linear Regression	20.81%	0%	20.5s
Improved ELM	12.84%	5.71%	10.5s

- The optimal hyperparameter value for linear regression is between $1e-9$ a $1e-6$.
- The optimal hyperparameter value for ELM is between 1 and 5 hidden neurons.

Al final de la presentación del póster, recibí el certificado de estancia de la NASA por parte de mi mentor el Dr. Martin.



El 13 de diciembre tuve una entrevista en la radio con el cónsul de México en San José, Carlos Ponce, donde se habló de la educación y las oportunidades académicas que ofrecía el gobierno mexicano.

En base al póster y con resultados nuevos, dí una presentación ante el departamento de sistemas inteligentes acerca de mi trabajo el miércoles 16 de diciembre.

La última presentación fue el 17 de diciembre ante trabajadores de la NASA de diferentes áreas.

Los resultados mostrados en el póster no fueron definitivos, y se continuo el trabajo hasta el último día, el 18 de diciembre. Con los resultados obtenidos del trabajo en conjunto se hizo una publicación científica del proyecto:

Aniruddha Basak, Jayasudha Jayakumaran, Ole Mengshoel, Mario Gurrola Morga, Stefan Hosein, Rodney Martin & Ishwari Aghav (2015). *Identifying Contributing Factors of Occupant Thermal Discomfort in a Smart Building*. Manuscrito publicado en Dashlink de NASA.

Mi vuelo de regreso a Guadalajara partió el 19 de diciembre, aunque debido a que se sobrevendieron boletos, regresé a México el 20 de diciembre.