UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Ingeniería



Diseño, Construcción y Caracterización de Detector de Radiación Vavilov-Cherenkov de Agua

Módulo:

Recepción, Almacenamiento y Análisis de Eventos Cósmicos con Distintos Algoritmos de Inteligencia Artificial

Trabajo de graduación en modalidad de Megaproyecto

presentado por

**Miguel Novella Linares**

Para optar al grado académico de Licenciado en Ingeniería en Ciencias de la Computación y Tecnología de la Información

Guatemala, Octubre 2016

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Ingeniería



Diseño, Construcción y Caracterización de Detector de Radiación Vavilov-Cherenkov de Agua

Trabajo de graduación en modalidad de Megaproyecto presentado por:

Daniel Eduardo Conde Villatoro

(Licenciatura en Física)

Pablo Roberto Duque Soto

(Ingeniería Electrónica)

Karen Roxana Guarcax Remaj

(Licenciatura en Química)

Miguel Novella Linares

(Ingeniería en C.C. de la Computación y Tecnologías de la Información)

Luis José Pinillos Motta

(Ingeniería Mecatrónica)

Gabriel Andrés Solomon Chajón

(Ingeniería Industrial)

Guatemala, Octubre 2016

Vo.Bo.:

(f) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Samuel Alfredo Chavez Fuentes

Tribunal Examinador:

(f)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ing. Samuel Alfredo Chavez Fuentes

(f)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ing. Luis Guillermo Mijangos Fuentes

(f)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Lic. Zaida del Rosario Urrutia

Fecha de aprobación: Guatemala, 8 de Noviembre de 2016

ÍNDICE

Página

Lista de Cuadros…………………………………………………………………….v

Lista de Figuras……………………………………………………………………..vi

Resumen……….……………………………………………………………………vii

1. Introducción…………………………………………………………………..1
2. Objetivos……………………………………………………………………...2
3. Justificación……………………………………………………………….….3
4. Marco Teórico…………………………………………………………….….5
5. Antecedentes en Física de Partículas y Rayos Cósmicos...…………....5
6. Implementación de un detector de Radiación Vavilov-Cherenkov en la Comunidad LAGO……………………………………………………………….8
7. Fundamentos de Inteligencia Artificial……………………………………11
8. Metodología…………………………………………………………………16
9. Resultados…………………………………………………………………..36
10. Discusión....…………………………………………………………………48
11. Conclusiones………………………………………………………………..54
12. Recomendaciones… ………………………………………………………55
13. Bibliografía………………… ……………………………………………….56
14. Anexos .……………………………..………………………………………58
15. Glosario……………………………………………………………………...59

LISTA DE CUADROS

Página

[Cuadro 1: Resultados del Entrenamiento del Clasificador Bayesiano 42](#_Toc465429474)

LISTA DE FIGURAS

Página

[Figura 1. Representación de una Cascada Hadrónica, con sus tres componentes identificadas 6](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946954)

[Figura 2. Comparación de dos partículas cargadas desplazándose por un medio a velocidad sublumínica (a) y superlumínica (b), ésta última perturbando el medio y emitiendo un cono de luz Cherenkov. 7](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946955)

[Figura 3. Componentes de un Detector Cherenkov de Agua 9](#_Toc465946956)

[Figura 4. Funcionamiento de un Tubo Fotomultiplicador (PMT) 10](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946957)

[Figura 5: Visita de Horacio Arnaldi a la UVG. De izquierda a derecha: Karen Guarcax, Luis Mijangos, Luis José Pinillos, Horacio Arnaldi (LAGO Bariloche), Miguel Novella, Daniel Conde y Pablo Duque. 19](#_Toc465946958)

[Figura 6: Diagrama de bloques de la electrónica LAGO 20](#_Toc465946959)

[Figura 7: Módulo de Presión y Temperatura y Módulo GPS utilizados 26](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946960)

[Figura 8: Interfaz del módulo GPS funcionando 27](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946961)

[Figura 9: Vista superior de una Red Pitaya 28](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946962)

[Figura 10: Prueba de modo Scaler de la Guía LAGO 30](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946963)

[Figura 11: Prueba de modo Histograma de la Guía LAGO 31](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946964)

[Figura 12: Prueba de modo Decaimiento de la Guía LAGO 32](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946965)

[Figura 13: Calibración de Eventos Cósmicos en sus tres componentes 33](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946966)

[Figura 14: Separación de Componentes de Rayos Cósmicos 34](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946967)

[Figura 15: Conteo de Eventos del 23 de agosto 37](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946968)

[Figura 16: Conteo de Eventos del 8 de octubre 37](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946969)

[Figura 17: Conteo de Eventos del 23 de septiembre 38](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946970)

[Figura 18: Histogramas de Amplitud y Carga del 23 de Agosto 39](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946971)

[Figura 19: Histogramas de Amplitud y Carga del 8 de octubre 39](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946972)

[Figura 20: Histogramas de Amplitud y Carga del 23 de septiembre 40](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946973)

[Figura 21: Histogramas de Carga en distintos intervalos del 23 de agosto 41](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946974)

[Figura 22: Conteo de tiempo entre Eventos consecutivos del 23 de agosto 42](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946975)

[Figura 23: Conteo de tiempo entre Eventos consecutivos del 8 de octubre 42](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946976)

[Figura 24: Conteo de tiempo entre Eventos consecutivos del 23 de septiembre 43](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946977)

[Figura 25: Modo Decaimiento en datos de Raspberry Pi 44](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946978)

[Figura 26: Modo Histograma en datos de Raspberry Pi 44](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946979)

[Figura 27: Modo Scaler en datos de Raspberry Pi 44](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946980)

[Figura 28: Clasificación Bayesiana aplicada a Eventos Cósmicos 45](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946981)

[Figura 29: Comparación de la Separación de Datos Simulados, Experimentales y de la Literatura 46](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946982)

[Figura 30: Resultado de Algoritmo KNN para K=3 47](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946983)

[Figura 31: Resultado de Algoritmo KNN para K=4 47](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946984)

[Figura 32: Error en Conteo de Eventos del 23 de agosto 49](file:///C:\Users\Anai\Desktop\Miguel\Semestre%2010\megaproyecto\Version%201.1%20Trabajo%20de%20Megaproyecto.docx#_Toc465946985)

RESUMEN

La Tierra está siendo irradiada en cada momento por una variedad de partículas provenientes del espacio exterior. Un tipo de partículas son los rayos cósmicos, los cuales provienen de lugares fuera del Sistema Solar. Esas partículas atraviesan la atmósfera y se descomponen en subproductos de otras, llegando eventualmente a la superficie.

Mediante la implementación de un Detector de Radiación Cherenkov en la superficie se puede captar éstos eventos. El objetivo de la presente fue implementar una versión de ese tipo de detector en la Universidad del Valle de Guatemala para capturar y analizar eventos cósmicos. Se buscó apegarse a los estándares de la comunidad LAGO (Latin American Giant Observatory), encargada de un proyecto de detección de cascadas de rayos cósmicos.

Como parte del módulo implementado se buscó, además de la construcción del detector, poder obtener y analizar los eventos capturados, mediante el uso de algoritmos de Inteligencia Artificial y Aprendizaje de Máquina (Machine Learning). Se realizaron procedimientos especificados por la Guía oficial de LAGO para asegurar la correcta implementación y como fundamento para el análisis y clasificación de los datos.

Se logró construir exitosamente un detector en la Universidad del Valle de Guatemala, así como se adquirieron los materiales para la construcción de un segundo detector con una mejor tecnología. Además, se logró clasificar e identificar correctamente los distintos componentes de los eventos resultantes de rayos cósmicos.

1. INTRODUCCIÓN

El Megaproyecto LAGO-UVG busca desarrollar e implementar un detector Cherenkov de Rayos Cósmicos, así como el manejo y uso de la información y datos generados por el mismo.

El diseño del detector consiste en un contenedor de agua, el cual será recubierto por dentro y modificado para contener un fotodetector. Dicho detector estará caracterizado y afinado para detectar Radiación Cherenkov correctamente en base al perfil electromagnético y otros factores de su ubicación en la UVG.

Como entregables del presente Módulo de Computación se incluye un diseño innovador de clasificación de eventos cósmicos, mediante el uso de Machine Learning y algoritmos de Inteligencia Artificial. El diseño permitirá diferenciar los eventos que acontecen, para su clasificación y posterior uso, o bien para ser descartados y ahorrar recursos. Se estará adquiriendo dicha información localmente, para que esté a disposición de la Universidad y sea de beneficio para sus estudiantes e investigadores (será de especial beneficio para los cursos de física avanzados, como apoyo para laboratorios y proyectos).

Además del almacenamiento local, se tiene contemplado transmitir los datos al repositorio de la comunidad LAGO para su uso por la comunidad científica Latinoamericana. Para dicha transmisión se requiere el acceso de parte de los miembros de la comunidad LAGO-Colombia al servidor utilizado, así como seguir el formato y procedimientos utilizados por la comunidad. Se siguieron las indicaciones del Ing. Horacio Arnaldi, miembro de la comunidad LAGO-Bariloche, para la construcción y prueba del detector a implementar.

1. OBJETIVOS
   1. OBJETIVO GENERAL
      1. Detectar, clasificar e interpretar Eventos Cósmicos de señales provenientes de un Detector Cherenkov de Agua.
   2. OBJETIVOS ESPECÍFICOS
      1. Utilizar Algoritmos de Inteligencia Artificial y Aprendizaje de Máquina para el Reconocimiento de Eventos Cósmicos.
      2. Calibrar el Detector implementado en base a los datos detectados, datos de la comunidad LAGO y la literatura correspondiente.
      3. Realizar algoritmos efectivos y eficientes que ejecuten las funciones del proyecto, especialmente en su rapidez e integridad de los datos.
      4. Enviar de datos de una Raspberry Pi a una Computadora local, permitiendo desplegar detectores en lugares remotos.
      5. Utilizar la Red Pitaya para adquisición de datos como alternativa de vanguardia.
2. JUSTIFICACIÓN

LAGO es una comunidad Latinoamericana que busca implementar una red no centralizada y colaborativa de Detectores Cherenkov de agua (WCD), integrada por investigadores y estudiantes de varias instituciones de los países miembros de la organización. Con ello buscan poder detectar y estudiar astropartículas de alta energía, con el fin estudiar la Climatología espacial, meteorología y otros temas relacionados. LAGO está conformada por varios WCD ubicados al nivel del suelo en distintos sitios de America Latina, cubriendo una amplia distribución latitudinal (lo que permite detectar cascadas de amplio alcance). Guatemala forma parte de la comunidad LAGO, pero no está transmitiendo datos actualmente. Es por ello que se debe resolver éste aspecto para aportar funcionalmente a la comunidad.

Existen antecedentes de proyectos realizados en la UVG de importancia para el proyecto LAGO-UVG. Entre ellos se incluyen los trabajos del Ing. Victor Araujo (respecto el uso de DSpace como repositorio de almacenamiento de eventos cósmicos de la comunidad LAGO) y el del Ing. Gustavo Castro (relacionado al uso de CORSIKA para simulación de cascadas de eventos cósmicos). Estos proyectos brindan algunos fundamentos necesarios para la implementación de un WCD conectado a la comunidad LAGO, por lo que es de importancia darles seguimiento y aplicar ese conocimiento en el presente proyecto.

Actualmente la comunidad LAGO está buscando mejorar y actualizar las herramientas y tecnología que utilizan en sus detectores. El equipo LAGO-UVG busca aprovechar esa oportunidad para innovar y mejorar aspectos como el hardware y programas utilizados. A pesar que el objetivo principal del proyecto es implementar un WCD utilizando la tecnología actual de la comunidad (una Nexys con una placa adquisidora) se estará realizando una versión adicional del Detector con una tecnología innovadora denominada RedPitaya. Dicha herramienta está especialmente diseñada para proyectos relacionados con Detectores de Radiación e investigaciones de Física de Partículas. La RedPitaya tiene varias ventajas sobre sus alternativas, ya que es un sistema todo-en-uno, pudiendo realizando las funciones equivalentes de una Nexys (procesamiento de los eventos y digitalización), una placa Receptora (para capturar los eventos análogos y transmitirlos a la Nexys) y una Raspberry Pi (la cual cuenta con un sistema operativo desde el cual se puede almacenar y analizar los datos digitalizados y también obtener parámetros importantes como la ubicación GPS, presión, temperatura, hora, etc.). No obstante, al ser una tecnología bastante reciente se deberá aprender a manejar y adaptarlo a las necesidades de la comunidad LAGO. Como equipo LAGO-UVG buscaremos realizar una comparación entre estas dos versiones de detectores, para así poder saber los beneficios de cada una.

El agua utilizada en los Detectores es de suma importancia, ya que sus propiedades lumínicas y de refracción afectan considerablemente la forma e intensidad de los eventos detectados. Además, es un medio dieléctrico dispersivo (radiador) que bajo ciertas condiciones emite Radiación de Vavilov-Cherenkov (que es capturada por el detector generando los eventos). El análisis del agua utilizada por la comunidad LAGO es un aspecto poco estudiado, a pesar de afectar los datos detectados y la calibración de los detectores. Es por ello que incluimos un estudio detallado del efecto del agua y sus propiedades, en la detección de eventos cósmicos.

Otro aspecto importante es que el detector permitirá a la UVG tener acceso a datos de primera mano de Eventos Cósmicos, pudiendo utilizarlos para motivos académicos en cursos avanzados de Física (como prácticas de Laboratorio, proyectos, etc.) y como caso de estudio para cursos pertinentes a cada una de las carreras de los integrantes (desde análisis químico del agua, aspectos teóricos de física de partículas, procesamiento y simulación de señales, análisis de algoritmos y clasificación de datos, etc.). Además, los datos disponibles pueden servir para la realización de variedad de artículos científicos, actividad que la Universidad debería realizar con mayor frecuencia tanto por los beneficios derivados del conocimiento generado (con una gran aplicación en el estudio de actividad Solar y su efecto en la Atmósfera) así como para un mejor perfil académico en general. Cabe mencionar que en evaluaciones recientes, la cantidad de artículos publicados por la Universidad del Valle de Guatemala fue el rubro que obtuvo la menor puntuación, representando un aspecto en el que se podría mejorar.

1. MARCO TEÓRICO
   1. ANTECEDENTES EN FISICA DE PARTÍCULAS Y RAYOS CÓSMICOS

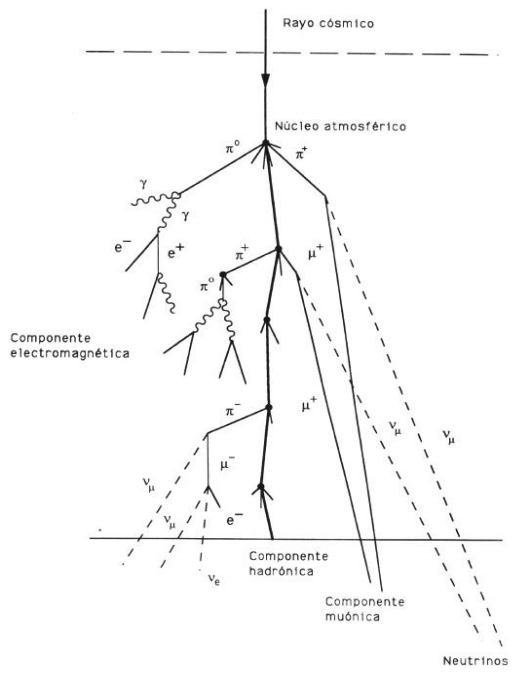
Los avances logrados el último siglo en Física de Partículas y Física Cuántica han permitido mejorar nuestra comprensión del Universo. Una de las explicaciones más prometedoras y detalladas de la naturaleza es el Modelo Estándar de Partículas Fundamentales, el cual es un desarrollo de la teoría cuántica de campos (unión de relatividad y física cuántica). Enormes avances han sido posibles gracias a los actuales aceleradores de partículas, permitiendo obtener y estudiar partículas de altas energías. (Asorey, H., 2012) No obstante, previo al desarrollo del primer ciclotrón (acelerador de partículas creado por E. Lawrence en 1932) la única fuente de partículas de alta energía capaces de producir nuevas partículas eran los rayos cósmicos. Dichos rayos fueron detectados por primera vez por V. Hess (1912) luego de una serie de experimentos a bordo de un globo aerostático, concluyendo que:

<<…una radiación de origen extraterrestre y de muy alto poder penetrante ingresa desde arriba a nuestra atmósfera. >>

Ese trabajo fue suficiente para valerle el Premio Nobel de Física en 1936 y además para que otros científicos realizaran experimentos similares con métodos más sensibles, dando inicio al campo de estudio de la Física de Partículas. Se define entonces los Rayos Cósmicos (Primarios) como las partículas, exceptuando los fotones, que se originan fuera del Sistema Solar y llegan a la Tierra y sus cercanías. (Asorey, H., 2012) Dichas partículas se originan por eventos altamente energéticos como supernovas, aunque también se pueden originar de estrellas (como el Sol) por llamaradas solares. (Mewaldt, R.A., 1996).

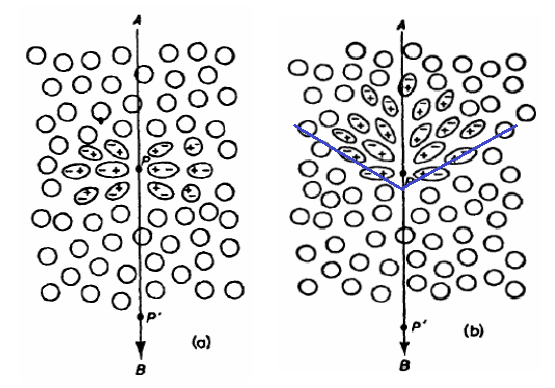
Posterior a su ingreso a la atmósfera, los Rayos Cósmicos continúan interactuando con los núcleos de los átomos de la atmósfera (como Oxígeno, Nitrógeno, etc.) produciendo hasta billones de partículas derivadas (Secundarias). A estas “cascadas” de partículas se les denomina Lluvias Atmosféricas Extendidas. Las lluvias atmosféricas extendidas poseen tres componentes principales: la cascada electromagnética, la cascada muónica y la cascada hadrónica. La evolución de una cascada en particular depende fuertemente de factores como la energía del Primario y la especie de rayo cósmico. Existen dos tipos de cascadas: las iniciadas por un fotón (γ) o electrón (*e*), aportando mayormente a las cascadas electromagnéticas (por reacciones como *Bremsstrahlung* y creación de pares de electrones) y las iniciadas por un hadrón (piones neutros: π0, piones cargados: π+ y mesones extraños como kaones: K+), que a su vez aportan a la cascada muónica. Éstos últimos son de mayor interés de estudio ya que contienen información relevante de lo ocurrido en las primeras interacciones de la cascada (sobre la partícula inicial), además de ser fácilmente detectables gracias a su comparativamente “larga” vida media. (Pérez, Y., 2015)

Figura 1. Representación de una Cascada Hadrónica, con sus tres componentes identificadas



El estudio de las Lluvias Atmosféricas Extendidas es el principal objetivo de la comunidad LAGO. Para captarlas, se busca implementar varios detectores ubicados a lo largo del continente americano que puedan captar eventos simultáneos producidos por esas cascadas de áreas extendidas. Los detectores que implementa LAGO son Detectores Cherenkov de Agua (WCD por sus siglas en inglés), capaces de percibir energía de la Radiación de Vavilov-Cherenkov (también referida solamente como radiación Cherenkov) causada por el paso de las partículas a través del agua del detector. La razón por la cual se detecta dicha energía, y por la cual se genera la Radiación de Cherenkov es debido a interacciones a nivel dipolo entre el agua y la partícula que ingresó en el tanque, la cual viaja a una velocidad mayor a la de la luz en su medio (agua). Cabe mencionarse que debido al índice de refracción del agua la velocidad de la luz en ese medio es del orden de 2.25 x 108 m/s versus 2.99 x108 m/s en el vacío. La radiación Cherenkov se diferencia en relación a la luminiscencia dado que tanto la polarización de las moléculas, la intensidad y tiempo de excitación no son alterables.  La luz emitida en este fenómeno presenta una asimetría notable, la cual se debe al movimiento de una partícula cargada a una velocidad de fase mayor a la de la luz en el medio. En donde por cada punto tocado por la partícula cargada se genera una excitación electromagnética, propagándose en la forma de una onda retardada desde estos puntos.  (Cherenkov, 1958) Éste fenómeno es comparable con su análogo acústico, en donde se generan ondas sonoras de choque cuando un objeto o partícula rompe la barrera del sonido.

Figura 2. Comparación de dos partículas cargadas desplazándose por un medio a velocidad sublumínica (a) y superlumínica (b), ésta última perturbando el medio y emitiendo un cono de luz Cherenkov.

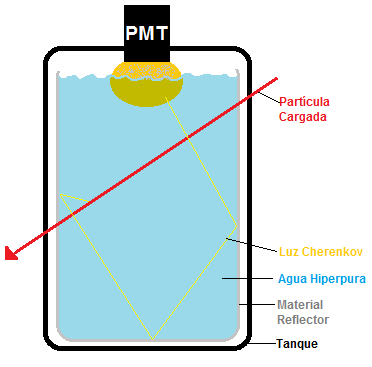


Las partículas detectadas por los WCD pueden ser electrones, muones que atraviesan (entran y salen) el tanque y además muones que entran al tanque y decaen (se degeneran en otras partículas) dentro del mismo. La forma de diferenciar entre los tipos de muones que son detectados es mediante la comparación de la energía que cada una de ellas emite en el agua, siendo dichas energías proporcionales al tamaño de la trayectoria dentro del tanque pero adicionando el denominado Michel Spectrum (energía ocasionada por el electrón resultante del decaimiento) en el caso de los muones que decaen. (P. Allison, 2005).

Como se mencionaba, el proyecto LAGO busca implementar varios WCD interconectados en un área extensa. En general, LAGO se ha dividido en dos estudios objetivos dependiendo de la altitud donde se encuentre el detector. Los detectores en altitudes superiores de 4,500 msnm permiten detectar Gamma-Ray Burst (GRB) con mayor sensibilidad; y los detectores de altitudes menores son utilizados para estudios físicos del sol llamados Forbush Decreases. Este  fenómeno consiste en la variación rápida de la intensidad de los rayos cósmicos, seguida por una eyección coronal desde el sol. Este efecto fue descubierto por Scott E. Forbush en 1937, que también observó que la intensidad de los RC al alcanzar la tierra se encontraban inversamente correlacionados con los 11 años de ciclo solar de las manchas solares (razón por la cual este tipo de proyectos tiene aplicaciones de monitoreo de clima solar).  (Villaseñor *et al.*, 2011)

* 1. IMPLEMENTACIÓN DE UN DETECTOR DE RADIACIÓN VAVILOV-CHERENKOV EN LA COMUNIDAD LAGO

La parte principal del presente Megaproyecto consiste en la implementación de un WCD propio de la Universidad, para poder adherirnos a la comunidad LAGO y comenzar a capturar eventos y transmitir datos. Un WCD consiste básicamente de un cilindro con paredes interiores reflectoras que contiene agua hiperpura como medio dieléctrico dispersivo y un tupo fotomultiplicador (o PMT por sus siglas en inglés). Cuando una partícula derivada de una Lluvia Atmosférica Extendida atraviesa el tanque, cumpliendo con las condiciones necesarias, se emite radiación de Cherenkov, la cual es recolectada por el PMT. Se puede notar que la cantidad de radiación emitida dependerá de la trayectoria de la partícula dentro del tanque (a mayor trayectoria mayor emisión). (De León, R., 2012)

Figura 3. Componentes de un Detector Cherenkov de Agua

Los detectores de la comunidad LAGO utilizan Tyvek como material reflector que recubre la pared interior del tanque. Tyvek es un polietileno de alta densidad con dos propiedades interesantes para la fabricación de WCD. La primera es su impermeabilidad al agua líquida, permitiendo contener el agua requerida por los detectores. La segunda, más importante, es que posee una baja absorción, permitiendo reflejar luz con facilidad. Ello permite que la radiación Cherenkov emitida por partículas que atraviesan el tanque sea reflejada y difundida en el interior del detector. Esto ocasiona que se forme una especie de “gas” de fotones dentro del tanque, aumentando la probabilidad de detección dado que ahora ésta depende solamente del área del detector. (Pérez, Y., 2012)

Los fotomultiplicadores (PMTs) son transductores lumínicos extremadamente sensitivos, con capacidad de detectar eventos de un fotón individual. Los PMT convierten fotones de longitudes de onda específicas en señales de corriente medibles. Posee una ganancia que lo hace ideal para detectar señales luminosas tenues y rápidas (como la radiación Cherenkov) y consiste de cuatro partes principales: Un fotocátodo, el cual libera un fotoelectrón al momento que un fotón colisiona con él (efecto fotoeléctrico); un sistema óptico basado en un arreglo de líneas equipotenciales para enfocar el fotoelectrón y poder dirigirlo; un multiplicador de electrones, el cual consiste de un arreglo de electrodos emisores que incrementan la cantidad de electrones generados (carga); y un ánodo que obtiene esa carga generada y la convierte en corriente medible. (De León, R., 2012)

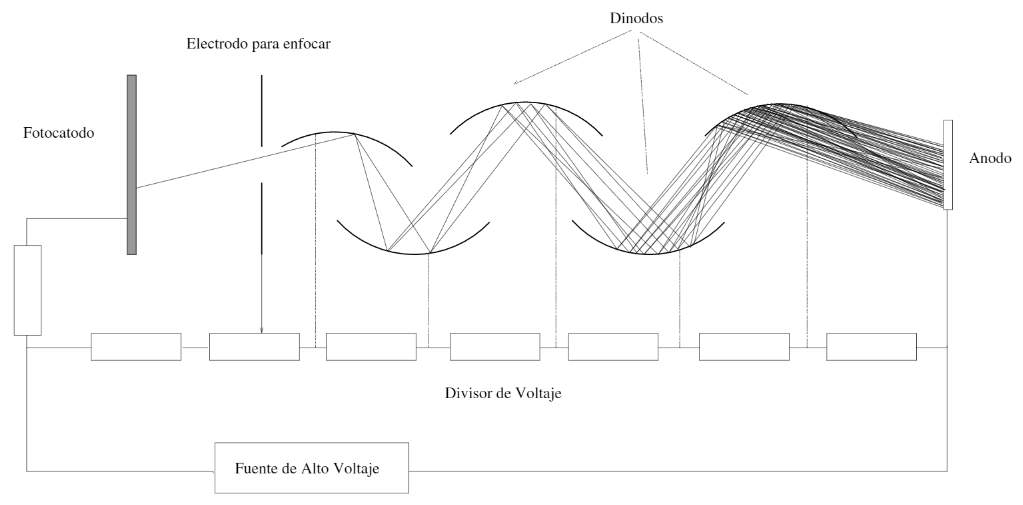


Figura 4. Funcionamiento de un Tubo Fotomultiplicador (PMT)

Las señales provenientes del PMT son analógicas, por lo que se deben discretizar (digitalizar) para su almacenamiento. La electrónica de LAGO utiliza conversores analógicos digitales de tipo flash (FADC en inglés), muestreando con una frecuencia de 40 MHz, para la digitalización de pulsos detectados. Con esa frecuencia específica se obtiene que el intervalo entre muestras (bin) equivale a 25 ns. Considerando que los eventos cósmicos tienen una duración promedio de 300 ns los datos digitalizados tendrán un tamaño de 12 bins. (Pérez, Y., 2012)

Una vez realizada la digitalización se obtiene una enorme cantidad de datos (34.56 Tb/día), de los cuales solamente unos pocos son de interés. Esto se debe a que los Rayos Cósmicos inciden al detector de manera aleatoria, por lo que la mayoría de los datos captados será ruido. Se desea entonces obtener sólo los datos que sobrepasen un límite de carga (trigger) específico, resultado de una emisión de luz Cherenkov debido al paso de una partícula. Para manejar volúmenes elevados de datos y en altas velocidades la comunidad LAGO utiliza dispositivos FPGA (Field Programmable Gate Array). Las FPGAs contienen bloques lógicos semiconductores que se pueden programar con lenguajes de descripción de hardware para configurar su uso, facilitando la discriminación de eventos recibidos por el PMT. (García, Luis G., 2015)

* 1. FUNDAMENTOS DE INTELIGENCIA ARTIFICIAL Y APRENDIZAJE DE MÁQUINA
     1. APRENDIZAJE SUPERVIZADO

La clasificación de los eventos registrados por el WCD se realizará mediante la implementación y entrenamiento de un Clasificador Bayesiano, mediante el algoritmo de Naïve Bayes complementado con técnicas de Alisamiento para mejores resultados.

El acercamiento en una clasificación Bayesiana es que se asume la independencia condicional entre eventos (de ahí su forma *naïve*), permitiendo calcular probabilidades de inferencia para elementos específicos.

La probabilidad para una categoría *C* dado un conjunto de atributos está dada por la siguiente expresión: (Russell, S. J., Norvig, P., 2010: 808)

( 1 )

En el caso de Naïve Bayes, el término de la multiplicatoria con Alisamiento de Laplace se calcula de la siguiente manera: (Thalmann, N.; Yuan, Q.: 2012)

( 2

Donde es el valor de Alisamiento, es la cantidad de veces que *x* aparece en el diccionario de la categoría *C* y |*V|* es el tamaño del vocabulario (todas las palabras). El alisamiento es algo sumamente necesario para el correcto funcionamiento del clasificador, dado le permite responder relativamente bien a información que no tenía clasificada previamente.

El clasificador Bayesiano se entrena de forma Supervisada, realizado distintas iteraciones hasta obtener un valor de alisamiento que minimice el error de los datos. Para realizar dicho entrenamiento se requiere obtener o generar un Lote de datos de entrenamiento que contenga información de varios eventos, los cuales deben estar clasificados (etiquetados) en base al tipo de partícula que representan (muones, etc). (Mitchell, T., 2015) Para obtener/generar dicho lote se consideraron las siguientes opciones:

1. Generación Manual de los Datos, etiquetados manualmente junto al resto del equipo.
2. Interpolación de Datos de otros miembros de la colaboración (Colombia) de X (x < 25) datos a uno con 25 valores, etiquetados manualmente junto el resto del equipo.
3. Combinación de 1) y 2)

De dicho lote de datos se realiza una partición en tres subconjuntos de datos: Datos de Entrenamiento (80%), Datos de Validación (10%), Datos de Prueba(10%). Dicha proporción es la recomendada para este tipo de algoritmos (MIT, 2015). Los datos de Entrenamiento serán los que se carguen inicialmente, para luego probar nuestro rendimiento con los datos de Validación. Esto se realiza iterativamente hasta obtener un resultado óptimo o se alcance el límite de iteraciones. Finalmente se utilizan los datos de Prueba para obtener el rendimiento final del clasificador.

Es sumamente importante estos datos de Prueba no sean incluidos en ninguno de los otros subconjuntos para evitar errores. Otro aspecto importante a considerar es que la totalidad de datos en el Lote deben reflejar la proporción verdadera de dichas partículas en nuestro universo, proporción que debe permanecer en los subconjuntos generados. Esto ya que el modelo responderá con más precisión a eventos que sigan esa proporción establecida en el entrenamiento. (Russell, S. J., Norvig, P., 2010)

* + 1. APRENDIZAJE NO SUPERVIZADO

Por lo general, los datos que se obtienen en distintos ámbitos científicos no poseen toda la información necesaria para calcular adecuadamente su probabilidad y definir un modelo. Las variables de la data puede dividirse en variables “ocultas” (aquellas propias del objeto de estudio pero no disponibles en un dato específico) y variables “visibles” (aquellas que sí conocemos). A las situaciones en las que no todas las variables son visibles en un conjunto de datos se les denomina Modelos de Variables Latentes; procesos importantes para la descripción del modelo pero que no se pueden medir directamente. Para éste tipo de datos se asume cumplen con el supuesto de Ausencia Aleatoria (*missing at random assumption*), la cual indica que los datos dependen solamente de los estados visibles (y no existe alguna condición que afecte de forma no aleatoria el modelo). Ello permite simplificar el proceso de aprendizaje pudiendo utilizar la probabilidad marginal para encontrar los parámetros optimizando la función:

( 3

Con respecto a los parámetros . Donde *v* corresponde a las variables visibles (los datos con los que se cuentan) y *h* corresponde a los estados de las variables ocultas.

Para maximizar esta probabilidad marginal se suele utilizar el algoritmo de Maximización de la Esperanza (*Expectancy Maximization*). El algoritmo EM es una forma iterativa y de propósito general para maximizar la probabilidad (esperanza) de un modelo de datos con variables ocultas o escondidas. Su implementación es sencilla y logra buenos resultados en pocas iteraciones. (Barber, David, 2012)

Formalmente, se puede obtener la siguiente cota para probabilidad marginal de un dato de entrenamiento individua *v*:

( 4

En donde *p(v, h | )* es el modelo paramétrico incluyendo las variables ocultas *h,* y *q(h | v)* es la distribución “variacional” de la probabilidad de que *v* pertenezca a *h.* El término variacional se refiere al hecho que la distribución mencionada es un parámetro de un problema de optimización. Sumando para todos los datos de entrenamiento se obtiene:

( 5

De ello deriva el algoritmo EM, que consiste de dos pasos: El **paso-E** encuentra las distribuciones que maximizan (5) con parámetros fijo y luego el **paso-M** encuentra los parámetros que maximizan (5) con distribuciones fijas. Dichos pasos se realizan de forma alternada e iterativa. Nótese la similitud de la segunda sumatoria con la ecuación (3); ello se debe a que se puede demostrar que basta con maximizar el segundo término de la ecuación (5) durante el paso-M dado que *q* no depende del nuevo parámetro obtenido. Además, la ecuación (5) resulta ser una cota exacta cuando , momento en que se obtienen los parámetros óptimos. Éste algoritmo garantiza un incremento en la esperanza con cada iteración. (Barber, David, 2012).

Otra manera de modelar datos de forma no supervisada es utilizar Modelos Mixtos (Mezclas). Un Modelo Mixto es un conjunto de modelos componentes que se combinan para producir un modelo más detallado:

( 5

Dónde *h* corresponde esta vez a los índices de cada modelo componente *p(v | h)* junto su peso o ponderación *p(h)*. Los Modelos Mixtos destacan en clasificar datos en conjuntos (*clusters*), en dónde *h* corresponde al índice del conjunto. Además, considerando *h* como una variable oculta los Modelos Mixtos pueden entrenarse utilizando el algoritmo EM. (Barber, David, 2012)

Los modelos Gaussianos son componentes convenientes a la hora de entrenar modelos mixtos, dado representan de forma continua probabilidades de masa ayudando a la interpretación intuitiva del modelo. Existen librerías de funciones especiales para trabajar con modelos mixtos, como la que facilita *scikit-learn* para Python (sklearn.cluster.Mixture). Una distribución Gaussiana D dimensional para una variable continua *x* es:

( 6

En dónde ***m*** es la media y ***S*** es la matriz de covarianza de tamaño DxD. Por lo que un Modelo Mixto de Gaussianos (GMM) tendría la siguiente forma:

( 7

La clasificación de los datos respecto al modelo mixto resulta de aplicar las consideraciones anteriores al algoritmo EM, resultando en un conjunto de parámetros para *i = 1,…,H.* Estos parámetros indican la media y covarianza de las *H* distribuciones Gaussianas que representan el modelo mixto. Un dato analizado con el modelo tendrá un porcentaje de pertenencia a cada uno de los conjuntos resultantes, clasificándolo como perteneciente al conjunto con la máxima probabilidad de pertenencia (argmax). Se puede notar que éste método brinda una clasificación “suave” respecto al modelo, otorgando un porcentaje de pertenencia a cada conjunto en vez de elegir una única categoría. (Barber, David, 2012)

Considérese ahora la siguiente Mezcla de *K* distribuciones isotrópicas Gaussianas , igualando sus covarianzas a la identidad :

( 8

Restringiendo el modelo a que todas las componentes tengas la misma matriz de covarianza la función de distribución de pertenencia se vuelve determinística (el dato *n* pertenece al conjunto *i* ssi. es el más cercano al valor de ese dato). Ahora los pasos E y M se realizan obteniendo el promedio de los puntos cercanos a . Éstas restricciones aplicadas al GMM lo reducen al algoritmo *k-means*, que a pesar de su simpleza converge rápidamente y brinda agrupamientos adecuados dado se tuvo una inicialización adecuada. El algoritmo *k-means* se utiliza usualmente como forma de compresión de datos, mediante cuantificación de vectores (*vector quantization*); reemplazando cada dato con el índice de su conjunto asociado. Es también utilizado para inicializar las condiciones iniciales de otros algoritmos de modelos mixtos, como el GMM. (Barber, David, 2012)

El algoritmo *k-means* construye *k* particiones del conjunto de datos minimizando la distancia entre cada elemento con el centro de cada uno de los *k* centros de conjuntos, de la forma:

( 9 )

Donde S es el conjunto de grupos a clasificar, *k* es la cantidad de conjuntos a clasificar, *x* es cada dato y *c* es el centro correspondiente a cada S. De esta manera se encuentra iterativamente el centro que minimiza el error o distancia entre datos, reflejando la mejor tendencia de los mismos. (Barber, David, 2006) Existen diversas implementaciones del algoritmo, como la que facilita la librería *scikit-learn* para Python (sklearn.cluster.KMeans).

1. MARCO METODOLÓGICO
   1. CONSTRUCCIÓN DEL DETECTOR Y ADQUISICIÓN DE DATOS

Durante la planeación y ejecución del presente Megaproyecto se siguieron dos metodologías: la metodología original, desarrollada por el equipo LAGO-UVG en base a la información inicial con la que se contaba, y la metodología final, desarrollada por la comunidad LAGO y adaptada por el equipo LAGO-UVG para el proyecto. Cabe mencionar el hecho que la metodología original resultó inviable por diversos factores que se mencionarán en breve. Se decidió seguir la metodología final dado ésta era más completa, se apegaba a los estándares de la comunidad LAGO y se adaptaba a los materiales y capacidades del equipo.

* + 1. METODOLOGÍA ORIGINAL

Una de las características de la comunidad LAGO es la implementación de varios detectores a lo largo del continente Americano, compensando la ausencia de un detector grande y en elevadas alturas con muchos detectores pequeños en un área extensa. Debido a estas características surgió la necesidad de construir detectores posiblemente en lugares remotos, con mayores alturas, para obtener mejores datos. Por lo tanto, se debe diseñar una alternativa de transmisión de datos desde el tanque detector hasta una computadora central. Para ello se optó por utilizar una Raspberry Pi junto el tanque detector encargada de adquirir los datos. El diseño del programa de Transferencia de Datos Raspberry-PC se dividió en dos partes: el programa ejecutado en la Raspberry Pi y el programa ejecutado en la computadora central. Se empleó un acercamiento de tipo Cliente/Servidor, en donde ambos utilizarán Network Sockets como forma de comunicación a distancia. El programa en la Raspberry se trabajó en lenguaje Python en el cual se implementará un Client Socket, quien abre conexiones bajo demanda y enviará paquetes de datos al Socket implementado en la PC. El programa en la PC se trabajó en el lenguaje Java. En esta parte se implementará un Server Socket quien recibirá constantemente los datos transmitidos por la Raspberry Pi y responderá un mensaje de confirmación de recepción correcta de los datos (momento en el que la Raspberry Pi dejará de retransmitir dicho paquete y procederá con el siguiente). (Kurose, J.F., 2012)

Posterior a la recepción, clasificación y análisis de los datos recibidos por el WCD se debe proceder a subir los datos al repositorio de la comunidad LAGO. El repositorio utilizado es DSpace, el cual se accede de una forma específica y configurando la máquina encargada adecuadamente. Trabajos previos indican ya se ha realizado este tipo de configuraciones, habiéndose realizado una Máquina Virtual configurada con los requerimientos para Acceso y Almacenamiento de datos en DSpace. Las características de la máquina que buscamos acceder son:

* + - 1. IP: 192.168.5.150
      2. S.O.: Debian 7 (64 bits)
      3. GNOME Desktop
      4. HDD: 109 Gb
      5. RAM: 4 Gb

Dicha máquina cuenta con las siguientes herramientas instaladas:

* + - 1. JDK 7
      2. Maven 2.2
      3. Ant 1.7
      4. Postgress 9 / Oracle
      5. Tomcat 7
      6. Entre otras… (Araujo, V., 2014)

En base a las consideraciones anteriores se plantearon varios requerimientos para los sistemas de Almacenamiento y Envío, Transferencia y Análisis correspondientes al presente módulo del Megaproyecto. No obstante, el cumplimiento de varios de los requisitos planteados resultó inviable, afectando la aplicación de la metodología inicial.

La transferencia de los datos capturados de la Raspberry Pi a la Computadora mediante envío de mensajes no fue posible debido a la enorme cantidad de eventos detectados (el tamaño promedio de cada hora de datos es del orden de 102 Mb). Transferir constantemente esa cantidad de datos mediante Wi-Fi (posiblemente desde un lugar alejado) hubiera resultado en una pérdida de paquetes elevada debido a la saturación de la red y las fluctuaciones de sus servicios. Se consideró como alternativa el envío de datos mediante GMS, dado presentaba ventajas sobre el Wi-Fi (como rango y disponibilidad). Se descartó dicha alternativa dado el costo monetario de transferir esa cantidad de datos era muy elevado.

El acceso a la Máquina Virtual mencionada en el trabajo de Victor Araujo no fue posible luego de varios intentos por obtenerla y recuperarla. Se comenzó por ubicar el servidor en donde residía dicha Máquina Virtual, identificándolo como el servidor ubicado en el aula J-205 de la Universidad (Laboratorio Avanzado de Computación). Para poder acceder al contenido del servidor fue necesario utilizar la el hipervisor de VMWare: VSphere. Con ésta herramienta se puede obtener acceso remoto a máquinas virtuales con facilidad. Una vez preparados los requerimientos para el acceso se encontró la dificultad de obtención de la contraseña y usuario del servidor. Para ello se consultó a varias personas del Departamento de Computación, miembros de la Asociación de Estudiantes de Computación e incluso exalumnos de la Unversidad que pertenecieron a la carrera de Computación y también a la Asociación de Estudiantes (la encargada del custodio de las credenciales). Luego de muchos intentos y consultas no se logró obtener la contraseña para el acceso a la máquina virtual y a ninguna parte del servidor del Laboratorio Avanzado. Se intentó además acceder físicamente al servidor, con la finalidad de recuperarlo, buscando obtener la contraseña o hash de la misma. Luego de varias pruebas de acceso y penetración no fue posible lograr entrar al servidor, por lo que finalmente se cesó el intento por acceder a la máquina virtual.

Los factores mencionados anteriormente afectaron seriamente el diseño del Megaproyecto, dado que una de las finalidades era unirnos a la comunidad LAGO aportando nuestros datos. Sin la Máquina Virtual como opción para envío de los datos se debería implementar una alternativa propia, o bien realizar nuevamente el trabajo de graduación de Victor Araujo (desviándonos de los objetivos del Megaproyecto). La metodología del Megaproyecto debió ser rediseñada drásticamente; afortunadamente se obtuvo ayuda de un miembro importante de la comunidad que permitió realizar e implementar el detector a tiempo.

* + 1. METODOLOGÍA FINAL

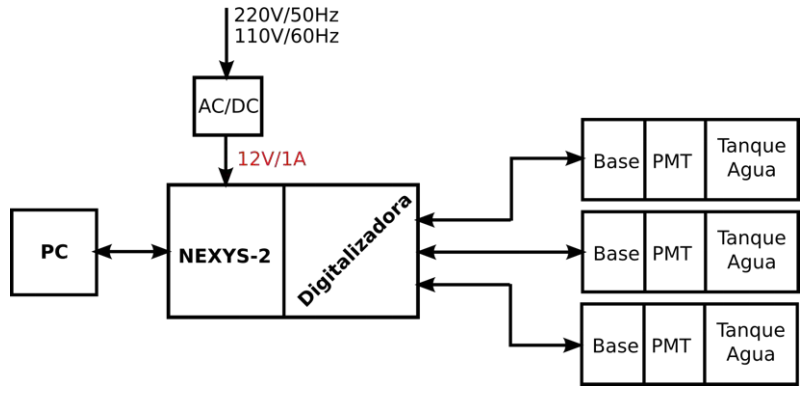
Para la semana entre el 18 y 25 de junio se contó con la visita de Horacio Arnaldi, Ingeniero Electrónico miembro de la comunidad LAGO de Bariloche, Argentina. Durante su visita impartió un curso a los miembros de la comunidad LAGO-UVG respecto a la implementación de un WCD. De estas clases impartidas es que se desarrolló la metodología final que se empleó en la construcción del Detector. Aprovechando la visita, el Ing. Arnaldi trajo consigo un PMT Photonis XP1802 y una placa digitalizadora donados por la comunidad LAGO; elementos necesarios para armar el detector. Además, se contó con la donación de una NEXYS 2 (otro elemento fundamental) de parte del Ing. Luis Arceo de la comunidad LAGO-México. Con los fondos recaudados por LAGO-UVG se compró un segundo fototubo de marca Hamamatsu R5912-20, designado para la construcción de un segundo detector con la nueva electrónica de LAGO (descrita en breve).

Figura 5: Visita de Horacio Arnaldi a la UVG. De izquierda a derecha: Karen Guarcax, Luis Mijangos, Luis José Pinillos, Horacio Arnaldi (LAGO Bariloche), Miguel Novella, Daniel Conde y Pablo Duque.



Como parte de la nueva metodología se siguieron los lineamientos especificados por la comunidad LAGO para el diseño de la electrónica del detector. La Base del PMT es el primero de estos componentes, y está encargada de recibir los pulsos provenientes del fototubo. Dichas bases, al igual que los fototubos de la comunidad LAGO, fueron rescatados del observatorio Pierre Auger. El componente más importante de las bases es la fuente EMCO capaz de producir una salida de 2000 V., el cual alimenta al PMT. Luego, los pulsos emitidos por el fototubo son transmitidos al siguiente componente: la Placa Digitalizadora. Esa placa puede digitalizar simultáneamente hasta tres canales de pulsos, pudiendo soportar esa misma cantidad de detectores. La digitalización la realiza con 10 bits a 40 millones de muestras por segundo (Msps), por lo que cada segundo se generan 50 Mb de información discretizada. Esa información es transmitida a la placa NEXYS-2: el kit de desarrollo de FPGA utilizado por la comunidad LAGO para la discriminación de pulsos. Esto se realiza dado que la mayoría de pulsos capturados son ruido, como se ha mencionado anteriormente. Por último, la transmisión de los eventos a la computadora receptora se hace mediante un conector USB mini. (Sofo Haro, Miguel, 2011)

Figura 6: Diagrama de bloques de la electrónica LAGO



Una vez digitalizados y discriminados los pulsos del PMT se puede proceder a adquirirlos y almacenarlos en la computadora receptora. El proyecto LAGO cuenta con un sistema de adquisición de datos integrados denominado ACQUA. Dicho sistema cuenta con los códigos de adquisición para las diferentes electrónicas de adquisición implementadas por la comunidad. Para la obtención de ACQUA se siguieron las especificaciones de la Wiki oficial del Proyecto LAGO, optando por utilizar un sistema Debian debido a su facilidad de configuración y para apegarse al estándar de la comunidad. Se inició instalando los requisitos necesarios para el correcto funcionamiento del paquete ACQUA:

* + - 1. Gcc completo
      2. Make
      3. Screen
      4. Small device C compiler cc1111
      5. Libusb

Mediante la ejecución del siguiente comando:

*sudo apt-get install build-essential screen libusb-dev cc1111*

El siguiente paso sería obtener la última versión del sistema ACQUA, mediante la clonación del repositorio oficial con el siguiente comando:

*sudo git clone* [*https://github.com/lagoproject/acqua.git*](https://github.com/lagoproject/acqua.git)

Una vez se ha descargado y descomprimido el paquete ACQUA es necesario realizar las configuraciones para la adquisición específicas para cada sitio. Para ello, LAGO ha adoptado un sistema de metadatos de autoingestión basados en el protocolo RSA de 2048 bits con una combinación de llaves pública y privada. Es necesario obtener primero un usuario y contraseña de parte del Responsable del Grupo de Datos, para poder transmitir los datos al repositorio central en Colombia. Dicho usuario obtenido cuenta únicamente con acceso a los protocolos SCP, SFTP y RSYNC, permitiendo copiar y transferir los datos y metadatos generados pero no ingresar ni navegar dentro del repositorio central.

Para facilitar y uniformizar la información generada ACQUA cuenta con el script *lago-configs.pl*, el cual realiza una serie de preguntas que sirven para configurar la adquisición en base a las características propias de cada detector. Al finalizar la configuración, se crea un archivo que contiene los aspectos ingresados. También se crea un archivo Crontab, encargado de ejecutar periódicamente los comandos necesarios para la adquisición, compresión, análisis y transferencia de datos. En la versión actual del sistema ACQUA las preguntas requeridas para la configuración son:

* + - 1. Datos del Sistema (Variable de entorno del directorio de adquisición)
      2. Datos del Sitio (Nombre, coordenadas y altitud, datos del responsable del sitio, institución y cantidad de detectores)
      3. Datos de los Detectores (Nombre, dimensiones, recubrimiento interno, modelo del PMT)
      4. Datos de Calibración (voltajes y niveles de trigger o disparo de cada detector)
      5. Datos de Adquisición (modelo de FPGA, modo de almacenamiento y transferencia)
      6. Datos de Sensores (GPS, presión y temperatura)
      7. Datos de Transferencia (transferencia automática habilitada, intervalo de transferencia, usuario en el repositorio, etc.)

Como se puede observar, en esta configuración se debe ingresar además el usuario y contraseña obtenidos para automatizar el proceso de almacenamiento de los datos. Estas variables son utilizadas extensamente por ACQUA, tanto al momento de la adquisición de eventos (voltaje, trigger, etc.) como referencia para los demás miembros de la comunidad que utilicen los datos generados (especificaciones del detector, ubicación, etc.).

Para finalizar, es necesario compilar el sistema de adquisición. Para ello, se aprovecha la configuración y variables de entorno ingresadas utilizando los siguientes comandos:

*cd $LAGO\_DAQ ; make*

Una vez finalizada la compilación se puede verificar la correcta instalación de ACQUA ejecutando el siguiente comando:

*./lago –v*

En este momento se debe desplegar información sobre la versión del sistema instalado (la versión más reciente al momento de la redacción del presente proyecto era la 5). Logrado esto la computadora ya se encuentra lista para adquirir y transferir datos capturados. Ahora el usuario puede interactuar con el sistema mediante el uso de los comandos *lago-start.sh* (el cual configura la NEXYS, modifica el crontab y verifica la correcta adquisición de los datos, volviendo a lanzarla en caso de ser necesario o posterior a una recuperación de una falla eléctrica) y *lago-stop.sh* (deteniendo la adquisición y modificando el crontab para detener la ejecución periódica de la adquisición). (LAGO Wiki, 2016)

Además de los dos comandos mencionados, ACQUA cuenta con otros más específicos que permiten realizar operaciones básicas para el manejo y obtención de datos. El procedimiento para encender el equipo es el siguiente:

* + - 1. Establecer el voltaje de la fuente en 12 V y una corriente de 1 Amp. Revisar la correcta conexión a los pines de la NEXYS (tierra y voltaje).
      2. Revisar que el PMT esté desconectado y la NEXYS con los jumpers como deben estar (posición WALL). Conectar la NEXYS a la computadora vía USB.
      3. Cargar el esquemático requerido por la NEXYS que la programa para la discriminación y recepción de pulsos. Debe tenerse cuidado de elegir el archivo correcto, ya que de ello dependerá lo que pueda hacer la NEXYS. Se utiliza el comando:

*./lago -x <archivo\_esquematico.xsvf>*

* + - 1. Revisar la conexión exitosa entre la FPGA y la computadora mediante los comandos:

*./lago –a* (estado de registros)

*./lago –t* (temperatura y presión)

*./lago –g (gps)*

* + - 1. Una vez se obtenga una conexión exitosa con la computadora se puede conectar el PMT a la placa digitalizadora. Además, se sugiere conectarlo a un osciloscopio para luego establecer un voltaje y poder observar los pulsos recibidos. Para indicar los voltajes y niveles de disparo (trigger) se utiliza el siguiente comando (reemplazando el nombre de registro por *hvX* y la cantidad por un número – v. gr. *1000*, donde X es el número de canal al cual se conectó el detector):

*./lago –s <nombre de registro> <cantidad>*

Para iniciar una adquisición de forma manual en modo de prueba basta con ejecutar el comando:

*./lago -f <nombre\_de\_archivo.dat>*

Lo que resultará en la creación de un archivo con el nombre especificado con todos los eventos recibidos y discriminados. Para detener la adquisición se debe terminar el proceso mediante el comando *CTRL + C* u otras alternativas.

Finalmente, el procedimiento para apagar el detector y electrónica de forma adecuada es el siguiente:

* + - 1. Establecer el voltaje del canal correspondiente en 0 V mediante el comando:

*./lago -s hvX 0*

* + - 1. Desconectar el PMT de la placa digitalizadora. Desconectar el conector de la base del PMT.
      2. Apagar la FPGA. Desconectar el cable de poder de la FPGA. Apagar la fuente de poder.

(Asorey, H.G., Gómez B., M., 2011)

La instalación del sistema ACQUA para recepción de datos también puede realizarse en una Rasberry Pi, permitiendo adquirir datos en un dispositivo de menor tamaño y poder implementar detectores en lugares alejados. Cabe mencionarse que la Raspberry Pi debe contar con acceso a internet, de lo contrario no será posible la transferencia de datos al repositorio y se podrán recuperar los datos únicamente accediendo de forma física el detector.

El sistema operativo de la Raspberry debe configurarse previo a su conexión con la FPGA y la recepción de pulsos. Se deben instalar los programas:

* + - 1. bash
      2. perl
      3. rsync
      4. ntp
      5. bzip2
      6. git (opcional pero recomendado)

ejecutando el comando:

*sudo apt-get install screen bzip2 rsync git ntp*

Luego se debe crear un usuario con nombre “lago” para uniformizar los distintos sitios de la comunidad. Además se le debe brindar accesos de superusuario con los siguientes comandos:

*sudo adduser lago*

*sudo visudo*

y agregar luego del usuario “pi” la línea:

*lago ALL=(ALL) NOPASSWD: ALL*

También se puede configurar el *daemon* NTP, para sincronización de la adquisición, y otorgar permisos a los dispositivos marca Diligent (el fabricante de la NEXYS 2) para evitar el uso de *root*. Dichos procedimientos se especifican en la Wiki Oficial del proyecto.

Una vez finalizadas las preparaciones de la Raspberry Pi se termina instalando el sistema ACQUA clonando el repositorio oficial, de igual manera a como se realiza en una PC. En caso de no poder instalar o compilar el sistema una alternativa es instalar las funcionalidades necesarias ejecutando el siguiente código:

*sudo apt-get update* (Actualizar la Raspberry y la fecha)

*sudo apt-get upgrade*

*sudo dpkg-reconfigure*

*sudo apt-get install g++ mercurial vim gnuplot-x11 / cc1111 libusv-dev libusb-1.0-0-dev screen* (Instalar dependencias)

*sudo chmod 777 –R /root/* (Otorgar permisos)

*cd /root/*

*sudo mkdir desarrollo* (Crear ambiente de desarrollo)

*cd desarrollo*

*sudo hg clone* [*https://lago@bitbucket.org/lago/libfpga*](https://lago@bitbucket.org/lago/libfpga)(Descargar)

*sudo hg clone* [*https://lago@bitbucket.org/lago/lago-daq*](https://lago@bitbucket.org/lago/lago-daq)

*sudo nano /home/pi/.bashrc* (escribiendo al final:)

*export MAKESTUFF=$HOME/desarrollo/libfpga/makestuff*

*source /home/pi/.bashrc* (Reiniciar bashrc)

*cd libfpga ; make* (Compilar ambas descargas)

*cd lago-daq ; make*

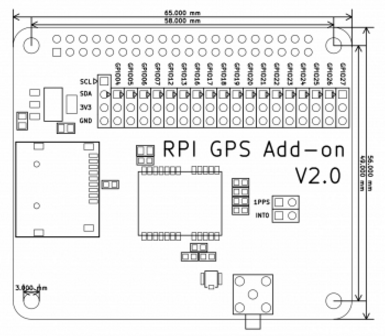
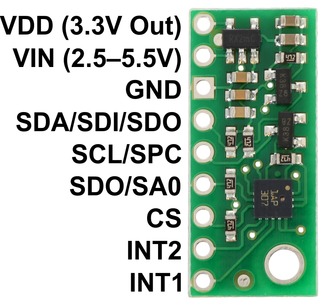
*sudo ./lago –h* (Probar correcto funcionamiento)

(Peña R., Jesús, 2012)

Los elementos instalados son las funcionalidades básicas del sistema ACQUA, consistiendo esencialmente del comando *lago* y sus respectivas opciones que permiten capturar pulsos y generar los archivos correspondientes. No se está instalando las funcionalidades de análisis de los datos, así como tampoco la automatización de los procesos de adquisición y transferencia mediante el uso de *lago-configs.pl* y los comandos *lago-start.sh* y *lago-stop.sh*.

La electrónica LAGO cuenta también con un sistema GPS y un módulo de Presión y Temperatura; datos necesarios para el almacenamiento y análisis de los pulsos capturados. No se pudo obtener los módulos utilizados por la comunidad LAGO, por lo que se optó como alternativa el uso de una Raspberry Pi para la implementación de ambos sistemas. Para ello se adquirieron dos componentes adicionales: El módulo *RPI GPS Add-on V2.0*, el cual se comunica mediante el puerto serial de la Raspberry Pi transmitiendo información como coordenadas, hora, etc. y el componente *LPS331AP MEMS pressure sensor* de STMicroelectronics, el cual brinda la presión y temperatura mediante el puerto serial con el protocolo I2C. Con estos componentes se procedió a diseñar un programa que se encargue de monitorear constantemente esos factores (coordenadas, presión, temperatura y hora) y posteriormente añadirlos a los archivos generados por ACQUA y también almacenarlos independientemente. Una vez se tengan modificados adecuadamente los archivos de pulsos capturados se puede proceder a transferir al repositorio central en Colombia.

Figura 7: Módulo de Presión y Temperatura y Módulo GPS utilizados



La implementación del sensor de Presión y Temperatura fue designada para el módulo de Electrónica (encargado: Pablo Duque). El diseño e implementación del módulo GPS fue designado al presente módulo, para el cual se siguieron los pasos basados en una implementación del mismo componente. (Causer, M., 2014)

Primero se preparó el puerto UART (serial) para el uso del módulo GPS, ya que por defecto éste es utilizado por Linux para iniciar sesión y depuración de kernel y puede contener tráfico de mensajes que afecten al componente. Para ello se deshabilitaron estas opciones modificando el archivo de inicio (boot) de la Raspberry Pi. Se recomienda obtener una copia de seguridad del sistema operativo antes de realizar las modificaciones, ya que un error puede ocasionar que el sistema no pueda iniciar nuevamente. Se aplicaron los siguientes comandos:

*sudo nano /boot/cmdline.txt*

modificando la línea:

*dwc\_otg.lpm\_enable=0 console=ttyAMA0,115200 kgdboc=ttyAMA0,115200 console=tty1 root=/dev/mmcblk0p2 rootfstype=ext4 elevator=deadline rootwait*

por lo siguiente:

*dwc\_otg.lpm\_enable=0 console=tty1 root=/dev/mmcblk0p2 rootfstype=ext4 elevator=deadline rootwait*

Se debe modificar también el archivo */etc/inittab* comentando la siguiente línea para evitar que se inicie sesión automáticamente:

*T0:23:respawn:/sbin/getty -L ttyAMA0 115200 vt100*

Finalmente se debe reiniciar la Raspberry Pi para que los cambios tomen efecto. Una vez realizado correctamente las configuraciones anteriores se continúa instalando el demonio GPSD (Global Positioning System Daemon), el cual se comunicará con el componente para obtener su información. Se debe instalar el programa junto sus dependencias, correr el servicio y probar su funcionamiento de la siguiente manera:

*sudo apt-get install gpsd gpsd-clients python-gps*

*sudo gpsd -F /var/run/gpsd.sock /dev/ttyAMA0*

*cgps -s*

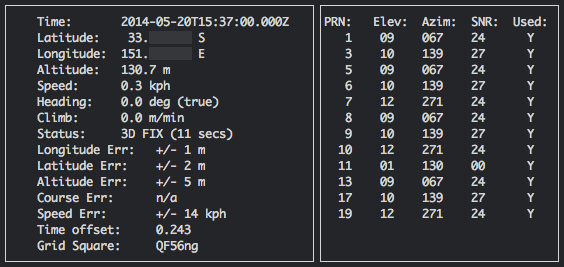
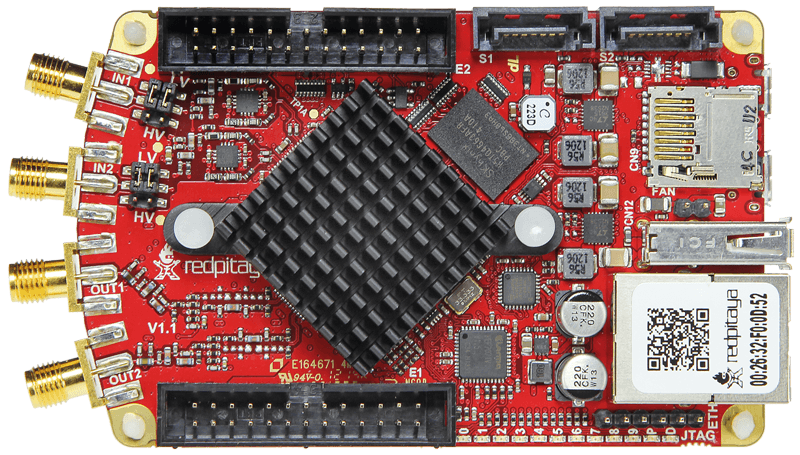
Si todo funcionó correctamente se deberá apreciar una interfaz similar a la siguiente:

Figura 8: Interfaz del módulo GPS funcionando

Para el diseño del programa que monitoree periódicamente los datos del GPS (Latitud, Longitud, Altitud, etc.) se decidió utilizar la funcionalidad *gpspipe* para consumir la salida del módulo. Dicha funcionalidad es parte del demonio GPSD y entrega paquetes de datos provenientes del módulo GPS mediante el uso de *pipes*. De esta forma se logra obtener una enorme flexibilidad ya que cualquier implementación (en cualquier lenguaje) puede obtener datos de *gpspipe* independientemente de la forma en que fuera diseñado el programa.

Como parte de los objetivos se buscó intentar implementar un WCD utilizando una Red Pitaya como componente electrónico encargado de la adquisición de pulsos. La Red Pitaya es una tecnología de vanguardia especialmente diseñada para proyectos que requieren rápidas velocidades de procesamiento e involucran grandes cantidades de datos. Cumple los roles equivalentes de la tarjeta digitalizadora, la tarjeta FPGA (como la NEXYS 2) y la computadora receptora en la electrónica de LAGO, reduciendo considerablemente la cantidad de componentes utilizados y el costo de obtener los mismos (tiene un costo aproximado de $300).

Figura 9: Vista superior de una Red Pitaya



La ventaja principal de utilizar la Red Pitaya es la mejora sustancial a la electrónica utilizada por la comunidad LAGO. Actualmente todas las versiones de electrónicas realizan la digitalización a 40 Msps con una resolución de 10 bits, como se mencionó anteriormente. La “desventaja” de estas velocidades es que se puede muestrear solamente en intervalos de 25 ns, hasta un máximo de 12 veces en un pulso que dure 300 ns (el promedio de los eventos). Ello ocasiona pérdida de información valiosa para la clasificación y análisis de los eventos detectados (Teorema de Nyquist-Shannon). La Red Pitaya tiene la capacidad de digitalizar a 125 Msps con una resolución de 14 bits, pudiendo muestrear en intervalos de 1/125\*106 = 8 nanosegundos. (Red Pitaya, 2015)

Esto se traduce en una mejora del 300% en la calidad de los pulsos capturados, lo cual beneficiaría enormemente el estudio de Rayos Cósmicos y sus aplicaciones. Es por ello que para el presente Megaproyecto se adquirió un segundo PMT, así como una Red Pitaya, anticipándose a la futura implementación de un segundo WCD con ésta nueva tecnología. Previo a la construcción del segundo detector se probará el funcionamiento de la Red Pitaya con los pulsos generados por el primer WCD implementado.

* + 1. CALIBRACIÓN, CLASIFICACIÓN Y ANÁLISIS DE EVENTOS CÓSMICOS

Los Eventos Cósmicos capturados y digitalizados por la electrónica de LAGO tienen la forma de pulsos discretizados en 12 intervalos (bins); cada intervalo corresponde a 25 ns debido a la velocidad de muestreo. Es necesario calibrar el detector y analizar los eventos antes de intentar clasificarlos con distintos algoritmos de Inteligencia Artificial y Aprendizaje de Máquina.

Para el Aprendizaje Supervisado se implementó un clasificador utilizando el algoritmo *Naïve Bayes*, el cual requiere etiquetar previamente los datos. Para ello se realizará la calibración del detector con la finalidad de poder clasificar los eventos en base a su carga para luego entrenar el clasificador.

En el caso del Aprendizaje No Supervisado se decidió utilizar el algoritmo de *k-means* y su variante el algoritmo *GMM-EM (Gaussian Mixture Model – Expectancy Maximization)*, ambos métodos de agrupación de datos (*clustering*). En este caso es necesario que las variables seleccionadas para la clasificación (*features*) presenten una distribución linealmente separable, razón por la cual se realizará un análisis visual de los eventos para determinar su aplicabilidad a dichos algoritmos.

* + - 1. Análisis de Eventos Cósmicos

Existen muchas formas posibles de analizar los pulsos capturados; la Guía de la Electrónica de LAGO realiza tres procedimientos iniciales que proveen visualizaciones de los eventos para su posterior calibración y clasificación.

El primero es el análisis en modo *Scaler*, el cual es un método de detección para detectores de superficie que determina las tasas de conteo de pulsos en intervalos de tiempo. Estas tasas son corregidas posteriormente en base a los registros de presión y temperatura de la ubicación del detector. Se utilizó el siguiente código mencionado en la Guía: (Arnaldi, H., *et al.*, 2001)

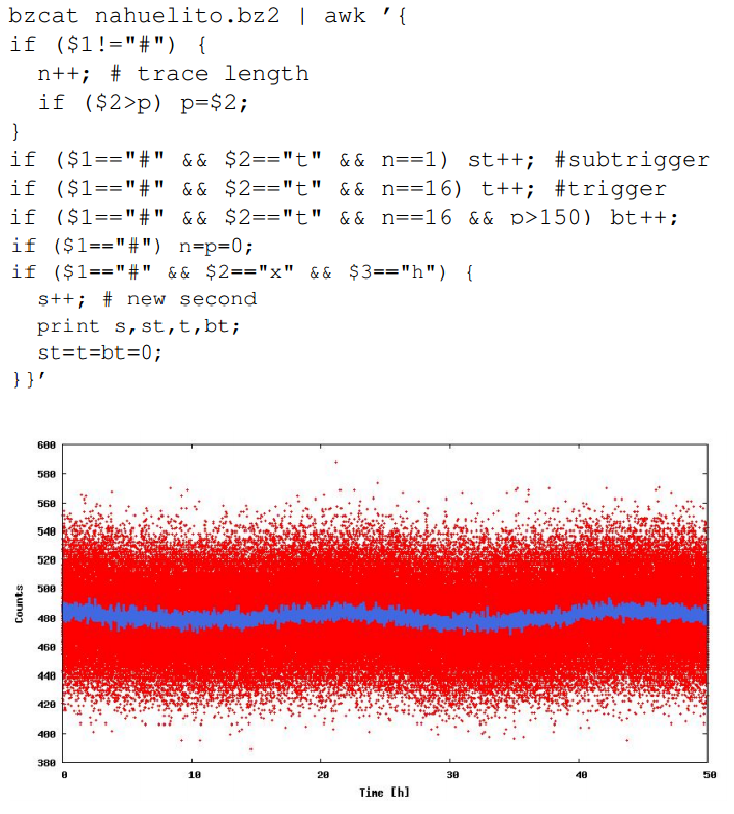
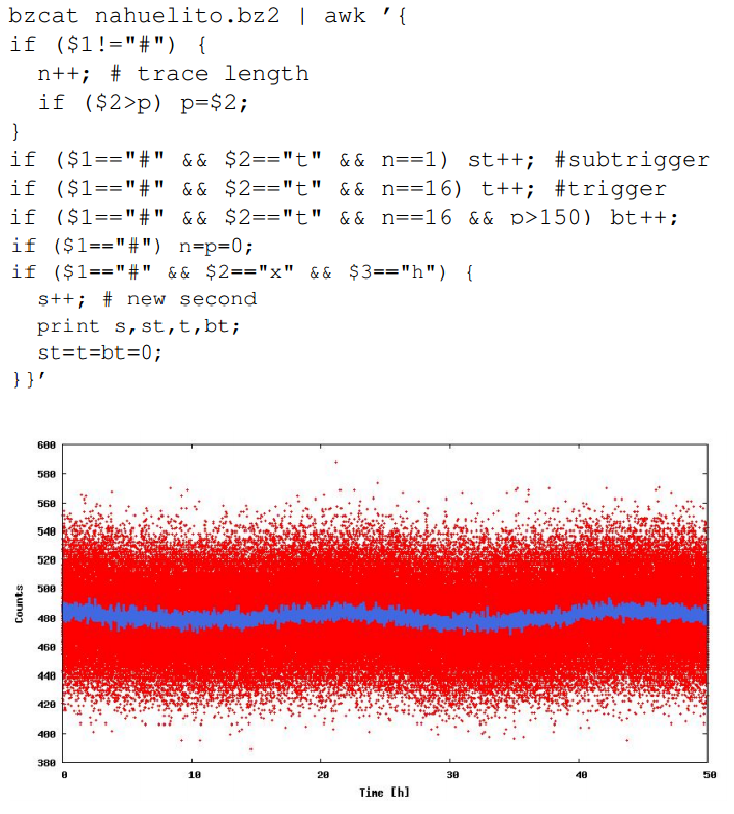


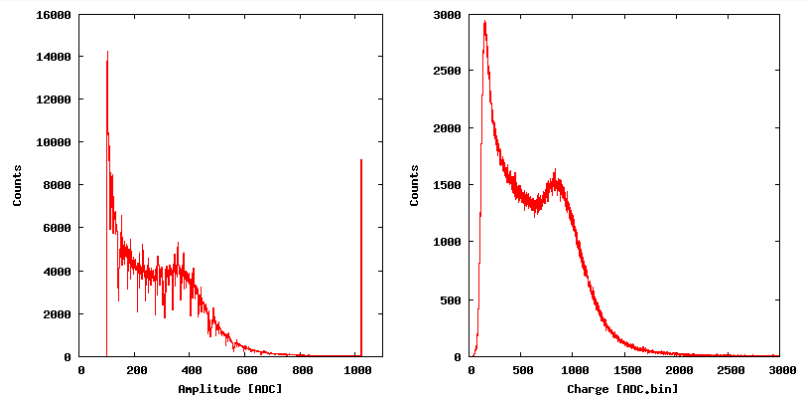
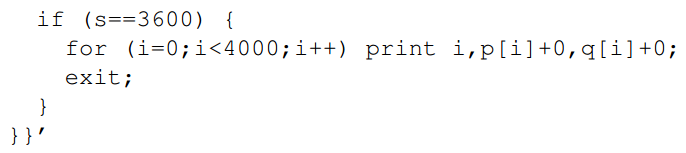
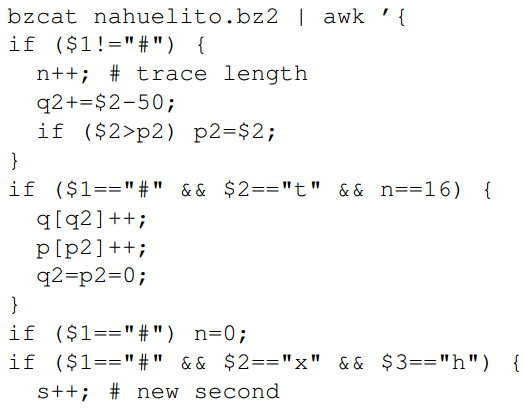
Figura 10: Prueba de modo Scaler de la Guía LAGO

El cual despliega el conteo de eventos en un intervalo de 50 horas, junto con el promedio por minuto indicado por la línea continua azul. El código permite también graficar subtriggers y triggers mayores (bt) de los datos.

La segunda forma es el análisis en modo Histograma, el cual permite el conteo de eventos en base a intervalos. Se realiza un conteo en base a la Amplitud (máximo ADC del pulso) y Carga (ADCq) de los datos capturados. Este modo es muy importante para la calibración del detector, dado que brinda las gráficas necesarias para determinar las cargas que separan las componentes Electrónicas, Muónicas y Hadrónicas de las Cascadas de Área Extendida. (Pérez, Y., 2015)

Para la obtención de los histogramas de la Amplitud y Carga se utilizó el siguiente código: (Arnaldi, H., *et al.*, 2001)

Figura 11: Prueba de modo Histograma de la Guía LAGO



Un tercer análisis que se puede realizar de los datos es la obtención de la vida de un muón (τμ), partiendo de la gráfica de diferencias de tiempo entre eventos consecutivos detectados. Para ello se seleccionan los eventos que cumplan con *C2 > C1*, donde *Ci* corresponde a la carga de segundo y primer evento. Dicha gráfica presenta una fuerte correlación con una función de decaimiento exponencial. Se utilizó el siguiente código para la obtención de la mencionada gráfica: (Arnaldi, H., *et al.*, 2001)

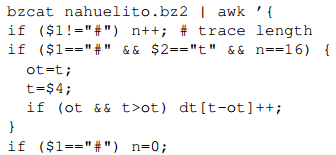
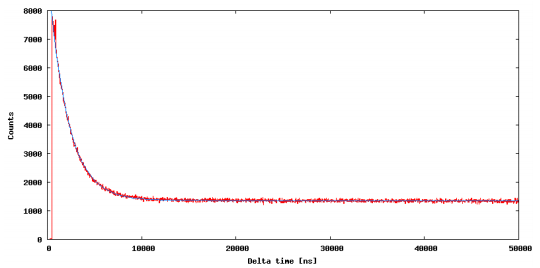


Figura 12: Prueba de modo Decaimiento de la Guía LAGO

* + - 1. Calibración de Eventos Cósmicos

La calibración del detector se realizó utilizando las gráficas generadas por los modos de Histograma (determinando la carga de las componentes Electromagnética, Muónicas y Hadrónicas) y Decaimiento (buscando encontrar la vida media del muón).

Para encontrar las cargas que corresponden a cada tipo de evento se siguió la metodología mencionada en un artículo. (Allison, P. *et al.*, 2005) Luego de obtener el histograma de carga siguiendo la metodología expuesta en la Figura 10 se realizó un ajuste gaussiano para encontrar el máximo del segundo pico de la gráfica. Este pico corresponde a muones que atraviesan el tanque de forma vertical (debido a que existe una mayor probabilidad que las partículas lo atraviesen de esa manera). Con ello se logra obtener la carga equivalente a un muón vertical en el tanque implementado (QVEM). También se analizó el mínimo entre los picos del histograma (Qem-max), correspondiente al fin de la componente Electromagnética e inicio de la componente Muónica.

Luego de obtener la medida QVEM para el WCD se siguieron los procedimientos de otro trabajo para obtener la carga máxima posible de un muón en base a la geometría del tanque. (Pérez, Y., 2015) Dicha carga se encuentra con la siguiente fórmula:

**( 3 )**

En donde *d* es el diámetro del tanque y *h*  es la altura del tanque. Con esta fórmula se obtienen los muones que atraviesan de forma diagonal el WCD (con mayor trayectoria, depositando más energía). Una vez obtenidos los parámetros Qem-max y Qμ-max se ha logrado diferenciar entre las componentes de rayos cósmicos; se puede notar que se cumple con Qem-max < QVEM < Qμ-max.

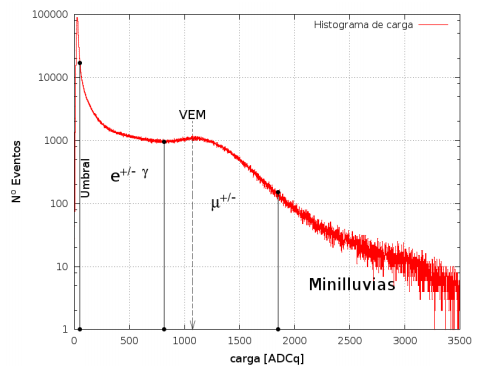


Figura 13: Calibración de Eventos Cósmicos en sus tres componentes

Se ha determinado teóricamente que la vida promedio de un muón (τμ) es de 2.19703 ± 0.00004 μs. (Muheim, Franz, s.f.) Utilizando los datos obtenidos en modo Decaimiento se siguieron los procedimientos encontrados en la literatura para intentar calcular experimentalmente τμ. Para ello se realizó un ajuste exponencial a los eventos graficados mediante , donde *b* son los eventos de fondo, *n* es la constante de normalización y *T* es el tiempo de vida del muón τμ. (Alarcón, M., 1999)

* + - 1. Clasificación de Eventos Cósmicos

Con los resultados de la Calibración del detector resulta trivial la generación de un conjunto de datos para el entrenamiento de un Clasificador Bayesiano. Los eventos de la componente Electromagnética corresponderán a aquellos con cargas Q < ; la componente Muónica serán los eventos con carga Q ≥ y Q < ; y las cascadas Hadrónicas corresponderán a los que posean carga Q > . Se clasificarán también aquellos eventos con carga Q ≈ como Muones Verticales al Tanque.

Los datos clasificados se dividen en conjuntos de Entrenamiento, Validación y Prueba, siguiendo las proporciones ideales mencionadas en la teoría (80%, 10% y 10%). Se procede a encontrar el factor que minimice el error de Prueba y se selecciona como parámetro para el Clasificador Final. Posterior a ello, cada nuevo evento capturado puede ser clasificado inmediatamente, permitiendo seleccionar aquellos que sean de interés para su almacenamiento definitivo y transferencia al repositorio central de LAGO.

Para la clasificación No Supervisada mediante Agrupamiento o *clustering* se debe iniciar seleccionando las variables que formen el espacio bidimensional sobre el cual se ejecutarán los algoritmos. Es importante seleccionar adecuadamente estas variables, para garantizar que los datos presenten cierto grado de separación natural que facilite su agrupamiento. Se optó por utilizar como principales variables la Amplitud (máximo) del pulso y su Tiempo de Subida, ambas utilizadas en un estudio similar que logra identificar las tres componentes de los rayos cósmicos. (Salazar, H., Villaseñor, L., 2005) Otras variables sugeridas por el estudio son ADCq versus Amplitud y Amplitud versus ADCq/Amplitud.

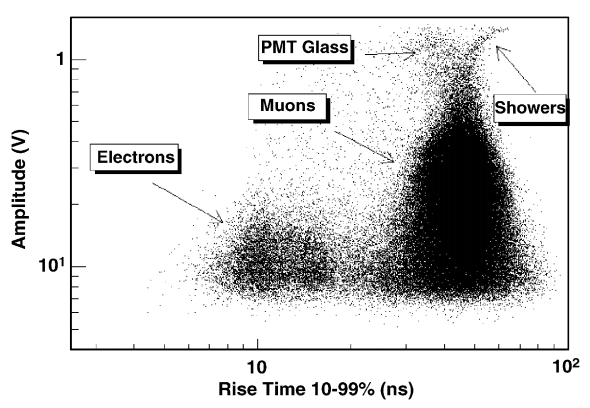


Figura 14: Separación de Componentes de Rayos Cósmicos

Se generaron archivos distintos con los eventos capturados para cada espacio bidimensional de variables. Se utilizó el lenguaje Python para la lectura de los eventos, debido a su facilidad en manejo de cadenas y texto y por su amplio uso para finalidades científicas (además para permitir la compatibilidad con la Raspberry Pi).

Para obtener el Tiempo de Subida (0-99%) se realizó un ajuste gaussiano a los pulsos capturados, mediante el uso de la librería *lmfit* para Python. Se obtuvo el tiempo que correspondía al máximo de la regresión mediante el comando *argmax* de la librería de análisis de datos *numpy*. Para asegurar que la escala de tiempo estuviera en nanosegundos se realizó la regresión con un espacio de datos de 300 elementos (dado la electrónica de LAGO utiliza 12 intervalos de 25 ns, resultando en una duración total de 300 ns).

La obtención de la Carga ADCq de los eventos se realizó integrando los 12 ADC que conforman el pulso. Para ello se utilizó la Integración por Trapezoide (*trapz)* proveída por la librería *numpy* para Python. La Amplitud se obtuvo de forma trivial seleccionando el elemento máximo entre los ADC del evento. En ambos casos es posible realizar la conversión de ADC a Voltaje siguiendo el factor de conversión de los PMT utilizados por la comunidad LAGO: (Pérez, Y., 2015)

**( 10 )**

Los 2V corresponden al rango de entrada del FADC implementado en la Tarjeta Digitalizadora y 210 corresponde a los 10 bits de resolución de la digitalización. Cabe mencionar que para la nueva Electrónica LAGO, utilizando la Red Pitaya, el valor de conversión será el siguiente (tomando en cuenta los 14 bits de resolución y el intervalo -1:1 de entrada FADC que utiliza):

**( 11 )**

Una vez generados los archivos están listos para aplicarles los algoritmos de Agrupación propuestos. Tanto el *k-means* como el *GMM-EM* fueron implementados utilizado la librería de Aprendizaje de Máquina para Python *scikit-learn*, debido a factores importantes como la eficiencia de los algoritmos, la documentación disponible y su compatibilidad con *numpy* y el resto de elementos del Megaproyecto.

1. RESULTADOS
   1. Construcción del Detector

La construcción del WCD para la comunidad LAGO-UVG se realizó siguiendo los lineamientos expuestos en la Metodología, con la finalidad de apegarse a los estándares y formatos utilizados por la comunidad en general. La construcción se realizó la semana entre el 18 y 25 de Junio en la cual se contó con la visita del Ing. Horacio Arnaldi de la comunidad LAGO-Bariloche, contando con la asistencia profesional de un conocedor del campo. A cada detector de la comunidad se le debe brindar un nombre con la finalidad de identificar la procedencia de los datos; el WCD implementado en la UVG se bautizó con el nombre de *Kinich Ahau*, haciendo alusión al Dios Maya del Sol.

Se inició con la preparación del tanque y el interior reflejante. El tanque consiste en un contenedor Rotoplast de 1500 litros, con 1.2 m de altura y 0.8 m de diámetro. Luego se realizó una estructura con tubos de PVC diseñada para colocarse dentro del tanque, funcionando como contorno sobre el cual envolver el cobertor fabricado con *Tyvek*. En paralelo se trabajó en la adaptación de la tapadera del tanque para poder contener el PMT y su base. Una vez terminados estas dos partes se procedió a ubicar el tanque en su posición final e iniciar su llenado con agua. Además se recubrió el tanque por fuera con un cobertor reflejante negro, para evitar el ingreso de luminosidad del ambiente que pudiera afectar la detección. Terminado esto el tanque se encuentra listo para iniciar su funcionamiento.

Como siguiente punto fue necesario instalar el sistema ACQUA para recepción de pulsos en una computadora designada para la tarea. El Departamento de Física designó una computadora del laboratorio C-112 para el uso por LAGO-UVG (Dell Precision T1700 con Ubuntu 14.02). Para la instalación de ACQUA se siguieron los procedimientos expuestos obtenidos de la Wiki y Guía Oficial de LAGO, tanto para la instalación en la computadora como para las pruebas realizadas en la Raspberry Pi.

Para las pruebas del presente Megaproyecto se realizaron varias tomas de eventos en las fechas: 23 de agosto (11 horas), 23 de septiembre (66 horas), 5 de octubre (3 horas) y 8 de octubre (21 horas). Los resultados más importantes son los correspondientes a los eventos del 23 de agosto y el 8 de octubre, ya que fueron las fechas en las que más tiempo se corrió el detector. La peculiaridad de ambas fechas es que cada una se realizó con distintos tipos de agua en el tanque (correspondiente al módulo de Química de Karen Guarcax); la primera con agua del servicio público y la segunda con agua tratada de forma especial. Con el agua tratada se buscaba eliminar la presencia de microorganismos que afecten los pulsos detectados sin modificar las propiedades de absorción de la misma. También se presentan los resultados probados con la Raspberry Pi (2 horas) realizado el 11 de agosto. Los demás eventos capturados se exponen en la Discusión en donde se analiza las posibles causas de los errores que afectaron a estos datos.

* 1. Modos de Detección de Eventos
     1. Modo Scaler para Conteo de Eventos

A continuación se presentan las gráficas de conteo de eventos para ambas fechas, basados en el modo de detección Scaler:

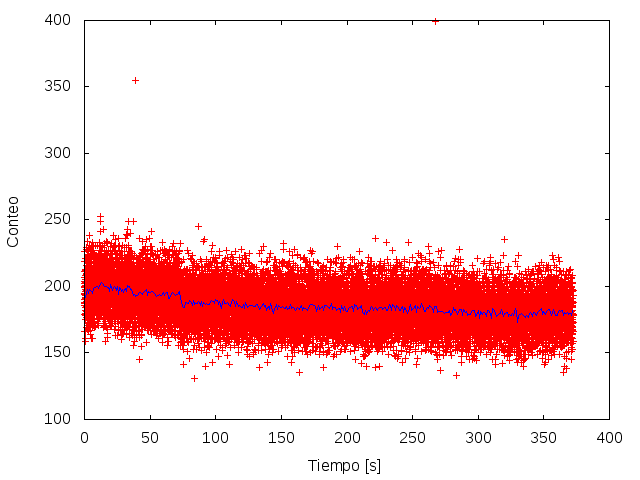


Figura 15: Conteo de Eventos del 23 de agosto

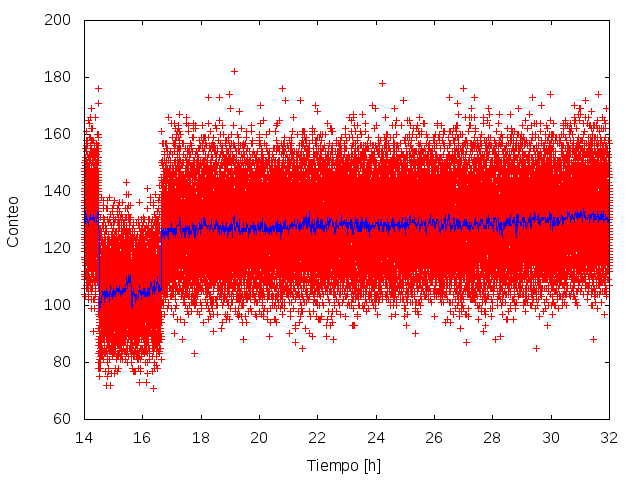


Figura 16: Conteo de Eventos del 8 de octubre

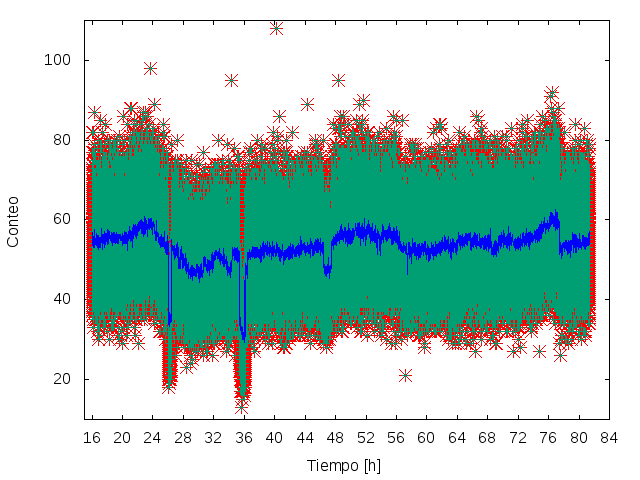


Figura 17: Conteo de Eventos del 23 de septiembre

* + 1. Modo Histograma para Calibración

Las siguientes gráficas corresponden a la Amplitud y Carga de los eventos para ambas fechas, basados en el modo de detección Histograma:

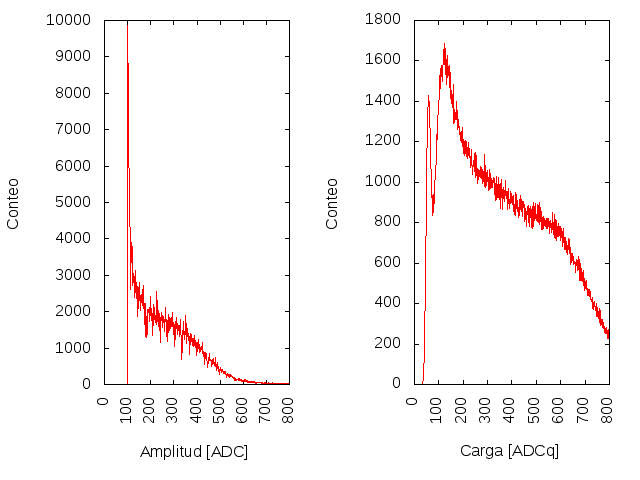


Figura 18: Histogramas de Amplitud y Carga del 23 de Agosto

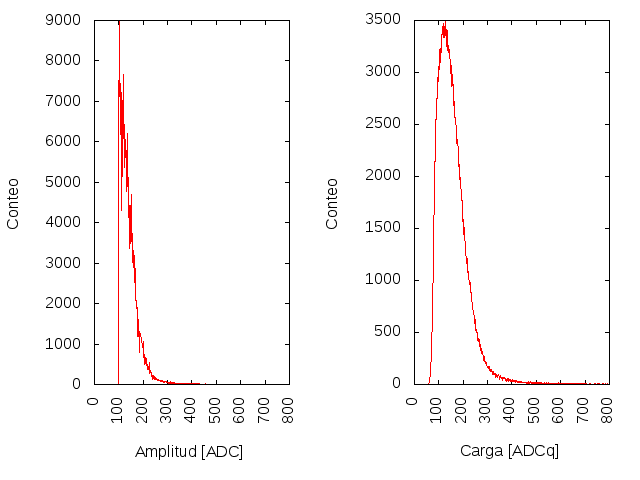


Figura 19: Histogramas de Amplitud y Carga del 8 de octubre

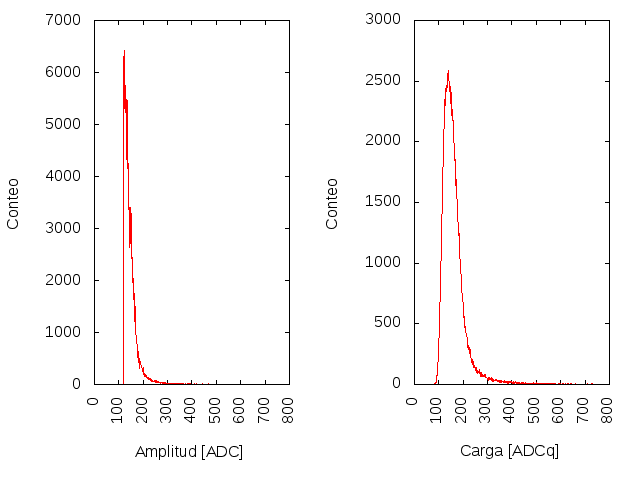
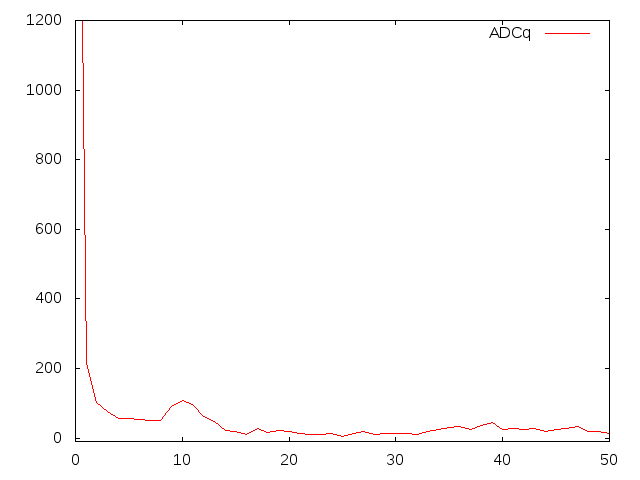
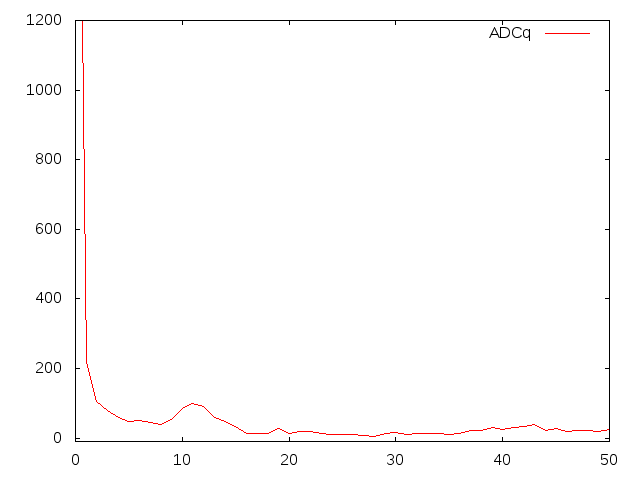
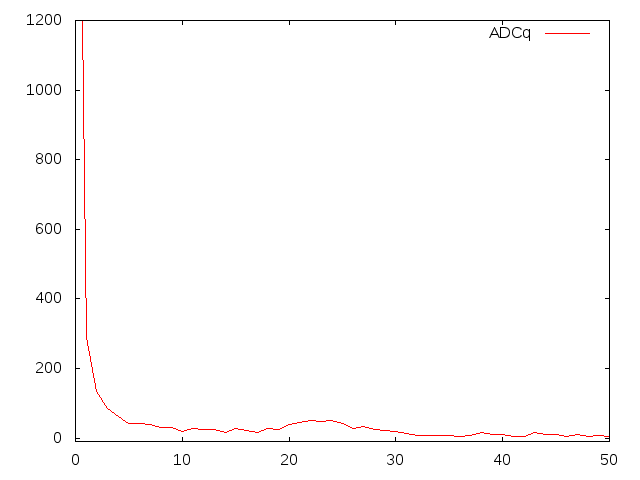
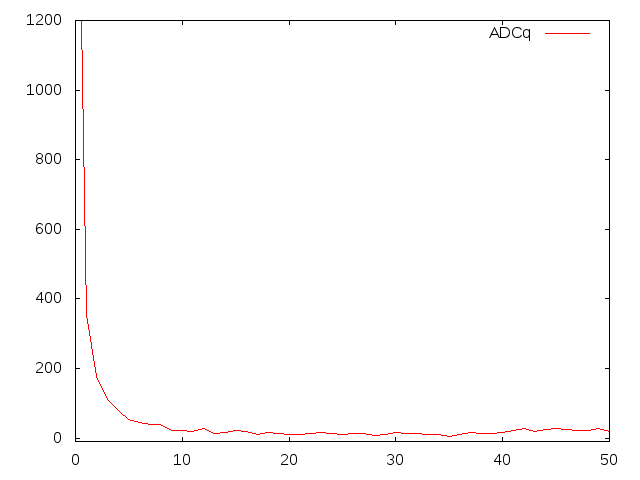


Figura 20: Histogramas de Amplitud y Carga del 23 de septiembre

* + 1. Histogramas obtenidos Manualmente

Adicional a los histogramas generados con el procedimiento indicado en la Guía Oficial de Lago se realizaron 4 histogramas similares, utilizando distintos intervalos de carga (en 5, 10, 22 y 25 nanosegundos). Ello con el objetivo de validar la carga máxima obtenida con distintas resoluciones de análisis. Los histogramas corresponden a los eventos del 23 de agosto.

Figura 21: Histogramas de Carga en distintos intervalos del 23 de agosto



* + 1. Modo Decaimiento para Vida de Muón

De similar forma se realizaron las gráficas de diferencia de tiempo entre eventos detectados, con la finalidad de encontrar el decaimiento exponencial que se ajuste a la vida media de un muón (2.197 microsegundos):

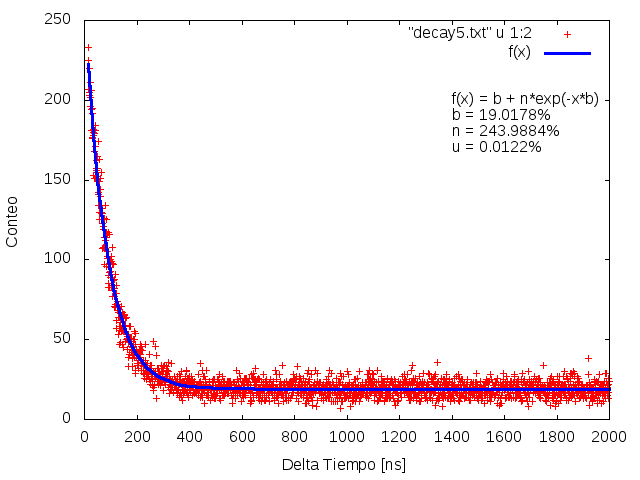


Figura 22: Conteo de tiempo entre Eventos consecutivos del 23 de agosto

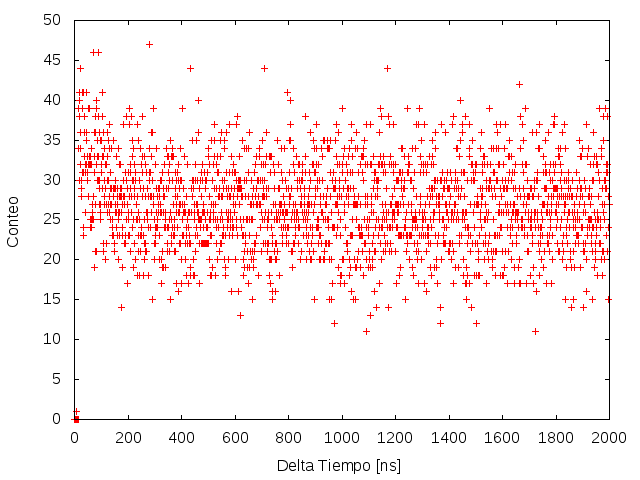


Figura 23: Conteo de tiempo entre Eventos consecutivos del 8 de octubre

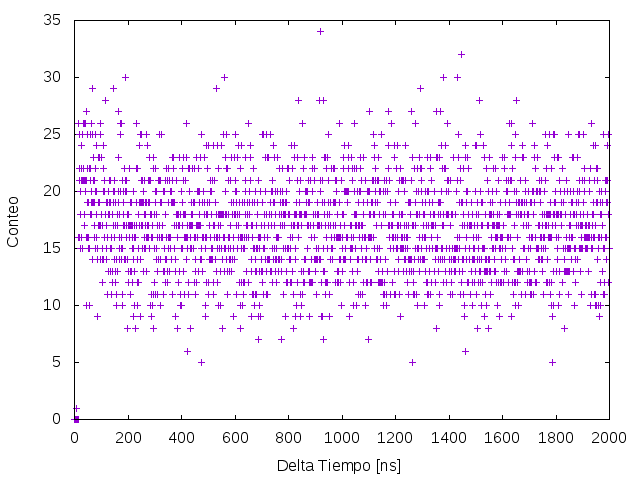


Figura 24: Conteo de tiempo entre Eventos consecutivos del 23 de septiembre

* + 1. Modos de Detección en Raspberry Pi

Las tres gráficas correspondientes a los modos de Detección para los datos capturados por la Raspberry Pi se presentan a continuación. Es importante mencionar que esos datos fueron tomados con el PMT fuera del agua (detectando luz Cherenkov emitida por partículas atravesando el aire).

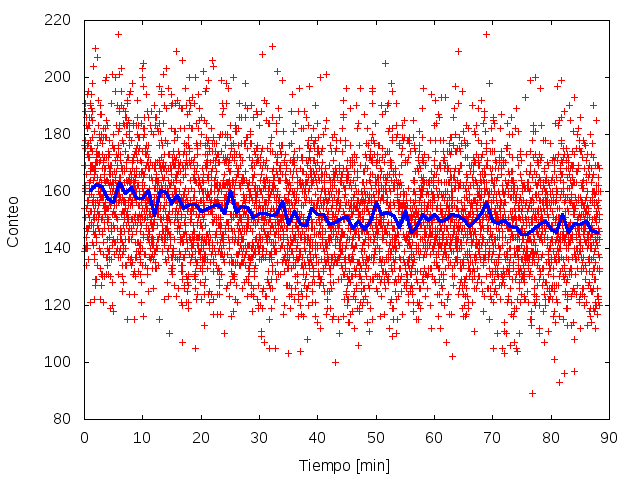
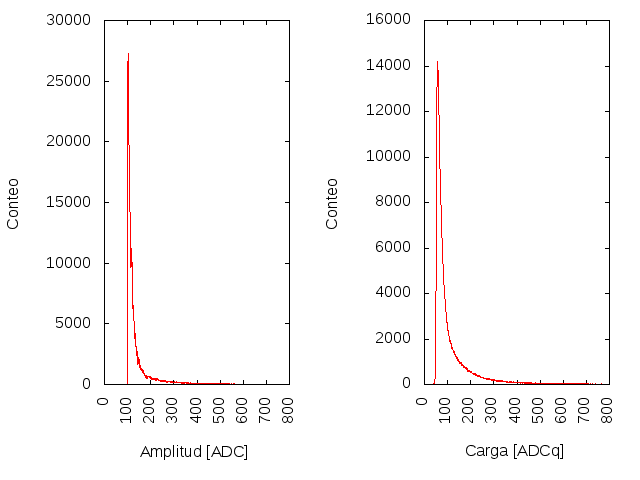
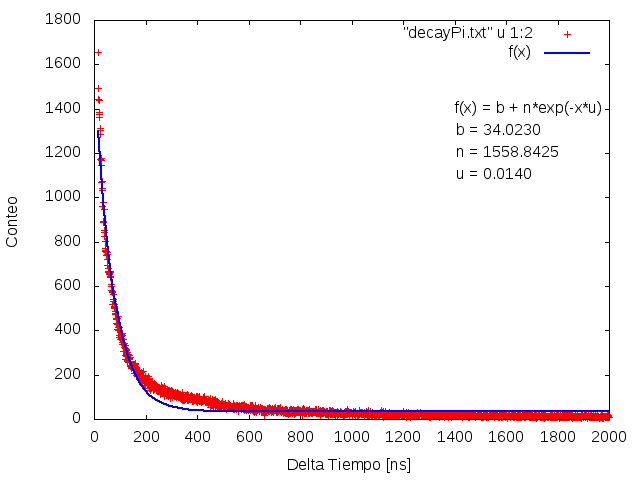


Figura 25: Modo Decaimiento en datos de Raspberry Pi

Figura 26: Modo Histograma en datos de Raspberry Pi

Figura 27: Modo Scaler en datos de Raspberry Pi

* 1. Clasificación de Eventos Cósmicos
     1. Clasificador Bayesiano con Alisamiento

A continuación se despliegan los resultados durante el entrenamiento y prueba del Clasificador Bayesiano. Luego, se presentan los datos clasificados según el resultado obtenido.

Cuadro 1: Resultados del Entrenamiento del Clasificador Bayesiano

|  |  |  |
| --- | --- | --- |
| **Etapa** | **Valor** | **Rendimiento (%)** |
| Validación | 0.5 | 0. 8599 |
| Validación | 10.0 | 0.8597 |
| Validación | 20.0 | 0.8595 |
| **Prueba Final** | **0.5** | **0.5456** |
| Validación 2 | 340.0 | 0.89023 |
| Validación 2 | 350.0 | 0.89030 |
| Validación 2 | 360.0 | 0.89017 |
| **Prueba Final 2** | **350.0** | **0.88605** |
| Validación 3 | 0 | 0.90061 |
| Validación 3 | 10.0 | 0.90556 |
| Validación 3 | 20.0 | 0.90262 |
| **Prueba Final 3** | **10.0** | **0.90079** |

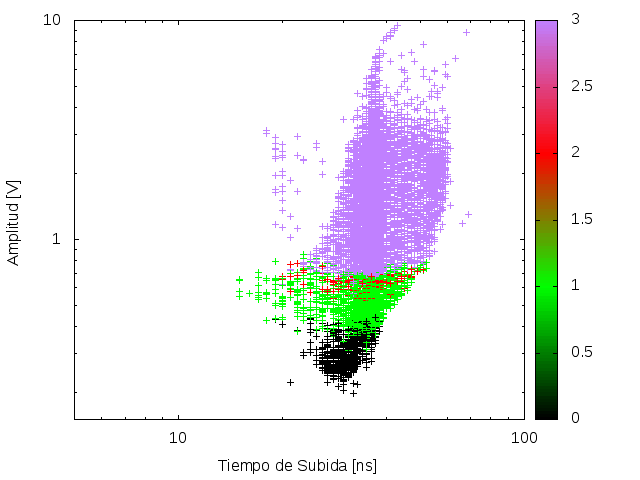
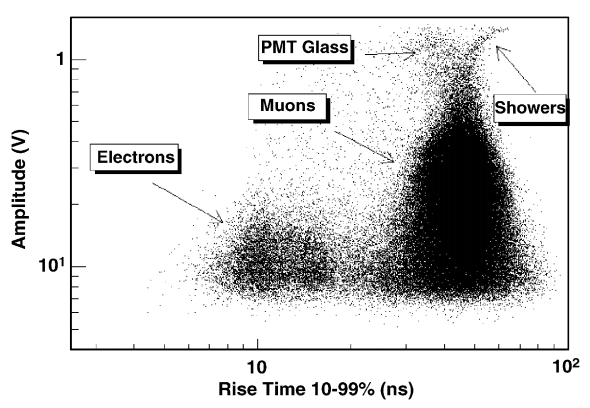
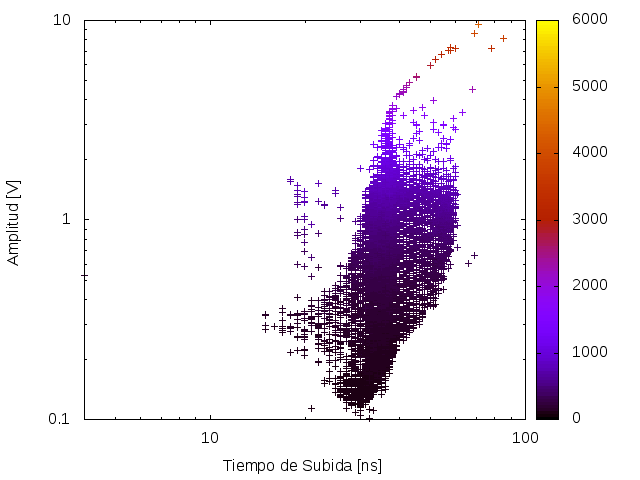
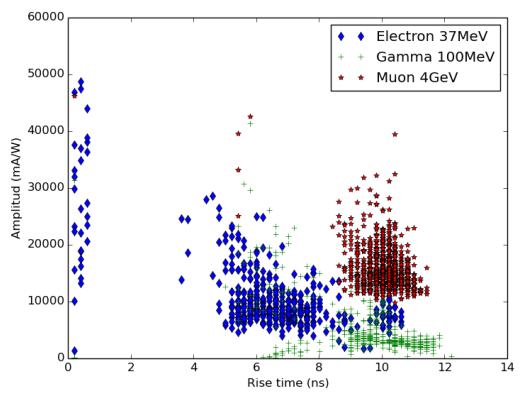


Figura 28: Clasificación Bayesiana aplicada a Eventos Cósmicos

* + 1. Clasificación No Supervisada

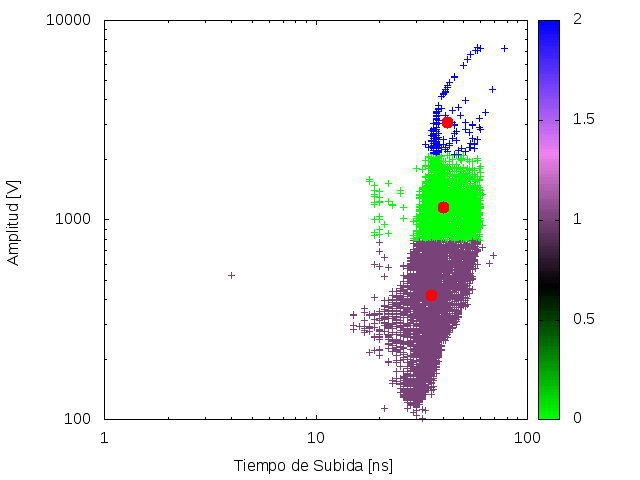
A continuación se muestran los resultados del intento de Separación de los pulsos mediante la comparación de su Amplitud y Tiempo de Subida. Se exponen los resultados obtenidos por la simulación realizada por Daniel Conde, los resultados obtenidos con los datos de agosto 23 y los resultados del artículo del cual se siguió la metodología:

Figura 29: Comparación de la Separación de Datos Simulados, Experimentales y de la Literatura



Se presenta a continuación el resultado del algoritmo knn aplicado a los datos separados del 23 de agosto (gráfica del medio en la Figura 27). Se intentó el algoritmo con *k = 3*  y con un *k = 4*:

Figura 30: Resultado de Algoritmo KNN para K=3



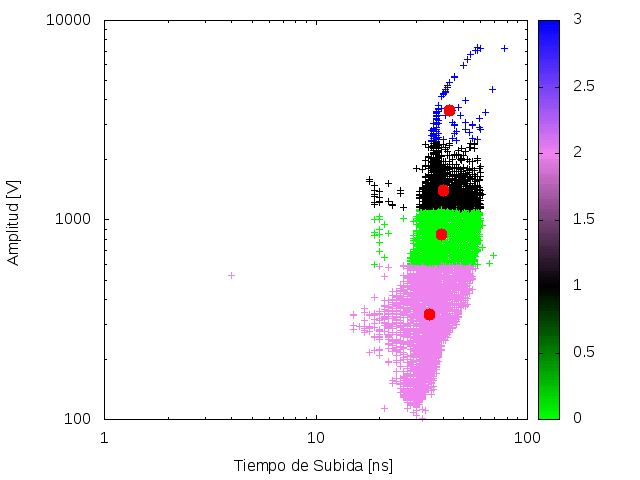


Figura 31: Resultado de Algoritmo KNN para K=4

1. DISCUSIÓN

La construcción del WCD fue probablemente la parte más trabajosa del presente megaproyecto. La obtención de los materiales requeridos fue compleja y costosa, ya que implicó en muchas ocasiones la compra en el extranjero de los mismos, como lo fue en el caso de la Raspberry Pi, el módulo GPS y de Presión y Temperatura y la tarjeta NEXYS 2. Incluso, el PMT Photonis del primer tanque fue traído por el Ing. Horacio Arnaldi, con sumas precauciones debido a la sensibilidad del instrumento. Al momento de utilizar el PMT Photonis por primera vez se encontró la fuente de alto voltaje de la base del fototubo se había quemado. Fue necesario adquirir una nueva fuente de emergencia para poder continuar con las pruebas.

La metodología a utilizar para la construcción fue otro aspecto difícil de decidir, dado al inicio no se contaba con el conocimiento de los estándares de la comunidad LAGO (por no contar con acceso a la Wiki y Guía oficial). Con la visita del Ing. Arnaldi se logró acceder a estas fuentes de información que permitieron definir una metodología apropiada.

La correcta implementación del detector se evidencia con la obtención de las gráficas para cada uno de los modos de detección, dado éstas presentan un fuerte parecido a los resultados esperados según la Guía. La obtención de las gráficas puede considerarse como una prueba de instalación exitosa del sistema ACQUA de adquisición de datos, cumpliendo con la parte de Detección del Objetivo General del Megaproyecto.

En los resultados se expusieron Figuras correspondientes a dos tomas de datos independientes: una con agua del servicio público y otra con agua tratada. El agua tratada contaba con un conjunto de químicos adicionados que garantizaban la ausencia de microorganismos que afectaran la detección de pulsos pero que no afectaban las propiedades físicas y químicas del agua que permiten la emisión de luz Cherenkov. Se esperaba obtener una mejor calidad de eventos detectados así como una constancia prolongada de los mismos (dado no habrá incremento de microorganismos con el paso del tiempo, afectando cada vez más a la detección en caso de no controlar dicho aspecto).

En las Figuras 14 y 15 se puede observar el conteo de eventos detectados en función del tiempo, para los datos del 23 de agosto (Corrida 1) y los datos del 8 de octubre (Corrida 4). Se puede notar un fuerte parecido entre ambas, sugiriendo que el agua tratada no afectó seriamente el conteo de eventos. En la Figura 14 se puede observar un decrecimiento ligeramente inusual cerca del minuto 75. Esta irregularidad se debió a que se removieron algunos datos capturados de la Corrida 1, dado se encontró un aumento extremo de los eventos capturados que afectaban el resultado del conteo. Se puede observar que luego de remover los datos erróneos (entre el minuto 100 y 200) la gráfica retiene su tendencia natural.

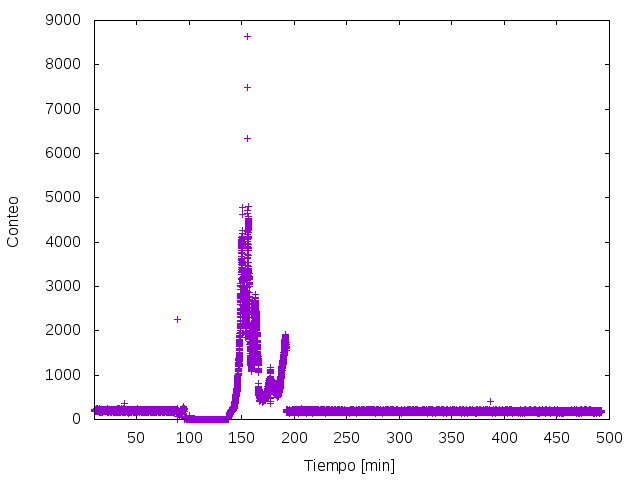


Figura 32: Error en Conteo de Eventos del 23 de agosto

En la Figura 31 se puede observar el conteo de datos totales de los eventos capturados esa fecha. Se observa un decremento significativo de eventos detectados, seguido de un incremento dramático y un período de inestabilidad. Momentos previos al minuto 200 (aproximadamente después de las 13:00h) la detección de eventos retornó a su conteo normal. Consultando con el equipo LAGO-UVG se logró deducir que se pudo deber a un incremento en la luz de ambiente que afecta el tanque, debido a una mayor incidencia de radiación solar de forma directa al tanque. Se reforzó la cobertura del tanque para evitar futuras fugas.

Observando detenidamente las Figuras 14, 15 y 16 se puede notar una diferencia interesante en la cantidad promedio de eventos detectados. El 23 de agosto se tiene un promedio cercano a 200; el 23 de septiembre se obtuvo un valor cercano a 60; y el 8 de octubre se obtuvo un valor aproximado de 130. El valor reducido obtenido en los datos del 23 de septiembre se atribuye a un error sistemático de la preparación del tanque para ese momento. Posterior a una inspección física se pudo observar que el recubrimiento Tyvek del tanque estaba dispuesto de manera incorrecta. Además, el nivel de agua en el tanque no era el suficiente para cubrir completamente el PMT. Estos factores ocasionaron una fuga de fotones, dado el Tyvek no pudo reflejarlos adecuadamente y además se contaba con menos nivel de agua, afectando los pulsos detectados, lo cual se tradujo en un conteo reducido de eventos. Se corrigieron éstos factores y se realizaron pruebas adicionales el 8 de octubre, momento en el cual se puede apreciar se obtuvo una mayor cantidad de eventos (130 promedio). A pesar de realizar las correcciones se puede notar que el conteo difiere a los obtenidos con el agua del servicio público (23 de agosto). La menor cantidad de eventos se puede atribuir al agua tratada que se utilizó en la Corrida de octubre, sugiriendo no cumplió a cabalidad su objetivo.

Se puede concluir algo similar observado las Figuras 17, 18 y 19 correspondientes al conteo de Amplitud y Carga de los eventos capturados. Nótese que en la Figura 17 se puede apreciar notablemente una forma similar a los resultados esperados de ejecutar el modo de detección Histograma de la Guía LAGO, en donde se puede observar los mínimos y máximos necesarios para la calibración del detector. No obstante, en las Figuras 18 y 19 no se aprecian éstos aspectos. Ello imposibilita la calibración del detector utilizando esos datos. Coincidentemente, las Figuras menos relevantes corresponden a las tomas de datos con el agua tratada.

Éste fenómeno se observa notoriamente en las Figuras 21, 22 y 23 del modo de detección Decaimiento, de las cuales sólo la 21 (correspondiente a los eventos del 23 de agosto con el agua del servicio público) presenta una tendencia exponencial apreciable. En las Figuras 22 y 23 resulta impráctico realizar un ajuste exponencial debido a la poca correlación que poseen los datos. Analizando el ajuste realizado a la Figura 21 resaltan los valores *b = 19.017* y *u = 0.0122*. El valor *b* (línea base o *baseline*) corresponde a ruido de fondo de eventos detectados. El valor *u* corresponde al factor de decaimiento exponencial, el cual aproxima la vida promedio de un muón. Comparando éste valor con la vida teórica del muón (2.197 microsegundos) con el valor *u* obtenido se puede notar que no logra aproximar de forma exacta dicho valor. Ello se debe probablemente a la escala utilizada para la regresión, pudiendo afectar al momento de hacer la regresión para *u* (propagación de error). También se le atribuye al hecho que se realizó el análisis de Decaimiento con todos los eventos detectados en esa corrida; realizar un filtrado previo seleccionando sólo los eventos muónicos (con ayuda de los algoritmos expuestos en la presente) se prevé ayudará considerablemente a la regresión obtenida. También realizar la regresión por otros medios puede ayudar a la exactitud de la misma. Otra fuente de error importante es la ubicación del WCD, el cual está dispuesto cerca del edificio C de la Universidad del Valle de Guatemala. Debido a la sensibilidad de los instrumentos es muy recomendable tener un espacio especialmente designado y considerado para ubicar éste tipo de instrumentos. Ello debido que factores como el ruido causado por electrones provenientes del decaimiento y desprendimiento del concreto de los edificios (situación expuesta en el trabajo de Salazar, H. y Villaseñor, L.). Se considera que de realizar las pruebas aquí expuestas en otra ubicación se podrán obtener resultados más limpios.

De manera similar se realizaron las pruebas capturando los eventos con la Raspberry Pi. Para esas pruebas se realizó la toma de pulsos sin sumergir el PMT en agua, simplemente realizándolas con el aire como medio asegurando no hubiera fuga de luz que pudiera afectar la prueba. Se puede observar en las Figuras 24, 25 y 26 los resultados obtenidos para los tres modos de detección utilizados. La Figura 26 presenta el conteo de eventos a lo largo del tiempo que duró la prueba, donde se puede observar un promedio cercano a 160 (con una tendencia decreciente). En base a la teoría se puede esperar ver un menor conteo de eventos detectados con el aire como medio. Ello se debe a las propiedades físicas del aire versus las del agua, teniendo ésta última una menor velocidad de fase para la luz (es decir, la luz viaja más rápido en el aire que en el agua). Con ello podemos concluir que una partícula deberá tener mayor energía para poder producir el efecto Cherenkov en el aire, lo que se traduce en una taza de detección menor debido a que existen menos eventos con tales energías. La Figura 25 presenta el histograma de Amplitud y Carga, pudiéndose observar que no cuenta con los factores importantes para la calibración del detector. Naturalmente, esto se debe también al medio utilizado para la detección. En el caso de modo Decaimiento, expuesto en la Figura 26, se aprecia un resultado aceptable; incluso muy similar al obtenido con el agua del servicio público. Finalmente se logró realizar una captura de datos correcta con la Raspberry Pi, mas no se optó por utilizarla activamente para las pruebas dado cuenta con menor capacidad de procesamiento que una computadora de laboratorio, pudiendo verse saturada por los datos de la Electrónica. Además, dado la Raspberry se conecta a internet inalámbricamente se debe contar con una señal bastante fuerte para poder garantizar que todos los eventos pudieran transferirse a algún lugar (como una computadora local o el Repositorio de Colombia). Es por ello que se requiere una mejor implementación de adquisición para Raspberry, en un lugar con condiciones más adecuadas y coordinado con el sensor de Presión y Temperatura.

Los factores mencionados anteriormente sugieren fuertemente que el agua tratada tuvo un efecto negativo imprevisto en la detección de eventos (se obtuvieron mejores resultados con el aire que con el agua tratada). Se considera que pudo ingresar contaminación al momento de llenar el tanque, previo a la ejecución de la captura de los datos. Ello pudo afectar bastante los resultados obtenidos en la calidad del agua, repercutiendo en la emisión de radiación Cherenkov que una partícula pueda dejar en su paso por el detector.

Respecto al resultado obtenido con el Clasificador Bayesiano se puede considerar se obtuvo resultados aceptables (Figura 27 y Cuadro 1). Se obtuvo un performance final del 56%, el cual puede no satisfacer inicialmente al lector, pero considerando que el conjunto de categorías en la clasificación Bayesiana estaba conformando por 4 elementos (Eventos Electromagnéticos, Eventos Muónicos, Eventos Muónicos Verticales y lluvias Hadrónicas) el resultado obtenido es por lo menos 100% mejor que la probabilidad de un performance totalmente *naïve* (adivinar con un 25% de éxito). Posterior a esos resultados se realizaron pruebas adicionales y se logró obtener un mejor resultado del 90% de performance final. La mejora se atribuye a una separación más rigurosa y adecuada de los eventos utilizados para los archivos de entrenamiento, validación y prueba (en las primeras pruebas no se siguió la proporción ideal de 80-10-10). No es de sorprender el hecho que se obtuvieran buenos resultados a pesar de los aspectos indeseados que pudiera presentar la data, dado esto es común debido a la versatilidad del acercamiento Bayesiano utilizado (según se ilustra en el libro de Russel y Norvig citado). Una mejora al método sería intentar otros tipos de alisamiento para mejorar el resultado, así como realizar varias mezclas del conjunto de datos para garantizar la independencia del entrenamiento.

Respecto al clustering de eventos se obtuvieron resultados muy interesantes, observables en las Figuras 29 y 30. Se puede notar que a pesar de que las pruebas de análisis de los datos indicaran visualmente que sólo se contaba con un conjunto identificable de datos el algoritmo logró clasificar una tendencia para *k=3* y *k=4*, notablemente lineal respecto a la Amplitud. Lo más destacable es que ésta tendencia lineal mostró similitud con los datos del Clasificador Bayesiano mencionado anteriormente, el cual también mostraba una tendencia lineal en su clasificación de los datos. Se recomienda fuertemente que se realicen más análisis de ésta variable de los eventos, así como incluir adicionales como Carga del evento (ADCq) o Delta Tiempo del evento anterior.

La Figura 28 presenta una comparación entre los resultados obtenidos por una simulación realizada en GEANT4 (C++) de los posibles eventos cósmicos incidentes en el tanque (realizado por Daniel Conde), los resultados experimentales obtenidos, y los resultados de separación que se encontraron en el trabajo citado de H.Salazar y L. Villaseñor (2005). Hay un fuerte parecido con los eventos de las Lluvias (nótese la cola en la parte superior de la Figura 24). La simulación se parece fuertemente al artículo, y presenta datos naturalmente separables. Los datos experimentales no presentan aparentemente la componente electrónica (con Tiempos de Subida alrededor de 10 ns). Éste factor se debió analizar detalladamente, concluyendo que se debió a la velocidad de muestreo utilizada por la electrónica (intervalos de 25 ns). Debido al teorema de Nyquist es que no es posible realizar una detección exacta de ésos eventos, dado se escapan de la velocidad de detección del WCD implementado. La nueva electrónica permitirá resolver ese problema al implementar una velocidad de muestreo superior, permitiendo obtener intervalos (bins) de 8 nanosegundos. Ésa cantidad es más adecuada para la detección del tipo de eventos mencionados (Correspondiente a la componente Electromagnética, debido al decaimiento de muones o bien por ruido de fondo). Los *features* seleccionados son adecuados para la clasificiación de los datos, mas se requiere de la nueva electrónica para que su aplicación sea más significativa (sólo se cuenta con un cluster visible por ahora).

No se incluyó en resultados pero sí se logró trabajar con la nueva Electrónica de LAGO (realizado en su mayoría por Luis José Pinillos). Se asistió en las pruebas con la RedPitaya, utilizando el lenguaje C++ para la utilización de las funcionalidades provistas por el API de la misma. En un futuro se buscará construir un segundo detector con la nueva tecnología exclusivamente para comparar ambas tecnologías. En las pruebas se logró recibir pulsos del mismo detector implementado, indicando la factibilidad de explorar esa tecnología. Se prevé trabajar en ello a futuro.

1. CONCLUSIONES

Se logró implementar exitosamente un detector de radiación Vavilov-Cherenkov de Agua apegándose a la tecnología y estándares utilizados por la comunidad LAGO.

Se calibró adecuadamente el WCD implementado, pudiendo diferenciar entre las distintas componentes de las cascadas de rayos cósmicos.

El Clasificador Bayesiano implementado en base al entrenamiento con datos calibrados tuvo resultados positivos, sugiriendo que mejores implementaciones son viables.

El algoritmo *k-means* implementado tuvo resultados positivos para *k=3* y *k=4*, mostrando una clara tendencia de agrupamiento lineal respecto la Amplitud de los pulsos. Aplicarlo a resultados de la nueva Electrónica brindará mejores resultados.

Los métodos de aprendizaje Supervisado y No Supervisado utilizados mostraron resultados parecidos, reforzando la validez de ésos algoritmos.

1. RECOMENDACIONES

Se sugiere fuertemente reubicar el detector a un lugar más adecuado. Las características deseadas de tal lugar son:

* + 1. Contar con espacio aéreo amplio, evitando estar rodeado por edificios o estructuras altas como montañas o paredes.
    2. Evitar el uso de material electrónico o radiactivo en las cercanías del detector (salvo los materiales del mismo).
    3. Fuerte ancho de banda de Internet, además de una fuerte conexión a internet inalámbrico o una alternativa similar.
    4. Seleccionar de preferencia un lugar con la mayor altura sobre el nivel del mar posible, para obtener mejor calidad de eventos cósmicos.

El presente Megaproyecto tiene muchas posibles aplicaciones y extensiones, por lo que se recomienda darle seguimiento. Las áreas en donde se puede mejorar lo implementado incluyen el diseño de una Electrónica más poderosa, la implementación de otros algoritmos para clasificación de eventos y el estudio de los mismos.

1. BIBLIOGRAFÍA

Alarcón, M., *et al.* 1999. *Calibration and monitoring of water Cherenkov detectors with stopping and crossing muons*. Nuclear Instruments & Methods in Physics Research, A 420 (1999). pp. 39-47.

Allison, P., *et al.*  2005.  *Observing muon decays in water Cherenkov detectors at Pierre Auger Observatory*.  29th International Cosmic Ray Conference Pune 00, 101-104.

Arnaldi, H., *et al.* 2011. *Data analysis of the high bandwith LAGO electronics*. Centro Atómico Bariloche, Argentina. pp. 8.

Asorey, Hernán, 2012. *Los Detectores Cherenkov del Observatorio Pierre Auger y su Aplicación al Estudio de Fondos de Radiación.* Tesis Doctoral en Física. Argentina: Comisión Nacional de Energía Atómica. pp. 360.

Asorey, H.G., Gómez Berisso, M. 2011. *New LAGO electronic: Getting started.* Centro Atómico Bariloche, Argentina. pp. 6.

Barber, David. 2006. *Machine Learning: A Probabilistic Approach*. 5ta Edición. MIT.

Barber, David. 2012. *Bayesian Reasoning and Machine Learning.* Cambridge University Press. pp. 697

Causer, M., 2014. *Raspberry Pi + ITead Studio GPS.* <https://github.com/mcauser/Raspberry-Pi-ITead-Studio-GPS-NEO-6M>. [20/6/2016]

Cherenkov, P. 1958. *Radiation of particles moving at a velocity exceeding that of light, and some of the possibilities for their use in experimental physics*. Nobel Lecture. pp. 15.

De León, R. 2012. *Detector Cherenkov de Agua*. Práctica Final en Licenciatura en Física. Universidad de San Carlos de Guatemala. pp. 30.

García, Luis G., 2015. *Diseño e Implementación de un Sistema de Adquisición de Datos para un Detector GRB en Guatemala*. Trabajo de Graduación, Facultad de Ingeniería, USAC. pp. 206.

Kurose, J. F., Ross, K. W. 2012. *Computer Networking*. 6ta edition. Pearson. <http://www.goodreads.com/work/best_book/80944-computer-networking-a-top-down-approach> [16/03/2016]

LAGO. *Wiki Oficial Del Latin American Giant Observatory.* <http://wiki.lagoproject.org> [18/10/2016]

Lawrence, E., Livingston, M. 1932. *The Production of High Speed Light Ions without the Use of High Voltages.* Physical Review, 40 (1), pp. 19-35

Mewaldt, R. A., 1996. *Cosmic Rays*. Macmillian Encyclopedia of Physics. <http://www.srl.caltech.edu/personnel/dick/cos_encyc.html> [11/7/2016]

Mitchell, Tom M. 2015.<<Chapter 3. Generative and Discriminative Classifiers: Naïve Bayes and Logistic Regression>>. *Machine Learning.* 5ta edición, McGraw Hill.

Muheim, Franz. *Muon Lifetime Measurement*. <http://www2.ph.ed.ac.uk/~muheim/teaching/projects/muon-lifetime.pdf> [26/10/2016]

Peña R., Jesús. 2015. *Conexión Raspberry Pi – LAGO.* Universidad Industrial de Santander, Colombia. pp. 4.

Pérez, Yunior, 2015. *Aplicación en meteorología especial de los datos del Proyecto LAGO (Latin American Giant Observatory).* Tesis Postgrado en Física. Mérida, Venezuela: Universidad de los Andes. pp. 84.

Red Pitaya. 2015. *Especificaciones de Hardware.* <http://redpitaya.com/#hw-specs> [26/10/2016]

Russell, S. J., Norvig, P. 2010. *Artificial Intelligence: A Modern Approach*. 3ra Edición, Prentice Hall. pp. 1132.

Salazar, H., Villaseñor, L. 2005. *Separation of cosmic-ray components in a single wáter Cherenkov detector*. Nuclear Instruments & Methods in Physics Research, A 553 (2005). pp. 295-298.

Sofo Haro, Miguel, 2011. *Electrónica LAGO: Guía de conexión de hardware.* Centro Atómico de Bariloche, Argentina. pp. 7.

Thalmann, N., Yuan, Q. 2012. *Enhacing Naïve Bayes with Various Smoothing Methods for Short Text Classification.* pp. 2

Villaseñor, L., *et al.* 2011. *Search for Gamma Ray Bursts and Forbush Decreases in LAGO Observatory*. ICRC 2011 Vol 10. pp. 314-317.

1. ANEXOS

El acceso a la Wiki de LAGO se puede obtener con las siguientes credenciales, brindadas por el Ing. Arnaldi para uso por la comunidad LAGO-UVG exclusivamente:

**Usuario**: lago

**Password**:

XsMqwb/BiI1UMuMvDPHrtbtBR69/DkkXulrOMNwhJNrBz2EKSrHameDTOZYf+Ix5

(Cifrado AES **llave**: rayoscosmicos)

1. GLOSARIO

Se presentan a continuación términos que podrían ser útiles para el lector en la comprensión de los fenómenos físicos y químicos involucrados en la detección de Rayos Cósmicos:

1. LAGO (Latin American Giant Observatory): Comunidad científica Latinoamericana que implementa un observatorio de rayos cósmicos mediante la construcción de un arreglo de detectores distribuidos en el continente Americano.
2. Modelo Estándar: Teoría Relativista de Campos Cuánticos (es decir, para altas velocidades y pequeñas masas) que describe la estructura fundamental de la materia y el vacío, considerando partículas elementales regidas por las cuatro fuerzas fundamentales conocidas.
3. Fermión: Uno de los dos tipos básicos de partículas del Modelo Estándar considerados los constituyentes básicos de la materia. Poseen espín semi-entero y se subdividen en quarks y leptones.
4. Bosón (de gauge): Uno de los dos tipos básicos de partículas del Modelo Estándar considerados los portadores de fuerza entre los Fermiones. Poseen espín entero e incluyen también partículas no fundamentales (compuestas) como el núcleo de deuterio.
5. Hadrón: Partícula subatómica conformada por quarks que permanece unida por la interacción nuclear fuerte. Ejemplos de hadrones incluyen neutrones, protones y piones.
6. Par de Electrones: Dos electrones que ocupan el mismo orbital molecular y por lo tanto poseen espines opuestos. Pueden encontrarse de forma “libre” (no como parte de una molécula) y son el resultado del decaimiento de un fotón, usualmente debido a un cambio de energía.
7. *Bremsstrahlung*: Radiación electromagnética (en forma de fotones) producida por la desaceleración de una partícula cargada (como un electrón). Aplica también a radiación debida a la aceleración de una partícula cargada, como la radiación de sincrotón o de ciclotrón.
8. Vida Media: El tiempo promedio en que un núcleo o partícula libre tarda en desintegrarse (en sus componentes fundamentales, como quarks). Representado con la letra tau (τ).
9. Rayos Cósmicos: Partículas fundamentales o compuestas (mayormente núcleos atómicos) que se originan fuera del Sistema Solar y llegan a la Tierra o su entorno cercano (denominados Primarios).
10. Lluvias Atmosféricas Extendidas: Posterior a su ingreso por la heliósfera, los Primarios interactúan con la atmósfera terrestre produciendo una enorme cantidad de partículas secundarias. Las lluvias atmosféricas son esos decaimientos producidos por la interacción de primarios con núcleos de elementos atmosféricos, los cuales eventualmente llegan a la tierra coincidentemente separados por áreas extensas.
11. Efecto Fotoeléctrico: Emisión de electrones por un material al incidir sobre él radiación electromagnética (fotones). La emisión se da cuando el fotón posee más energía que la función de trabajo del electrón, causando que éste último se separe del material.
12. ADC: Siglas de *Analog to Digital Counts*, o Cuentas de conversión Analógico-Digital. Es la unidad que representa la discretización de una señal contínua.
13. ADCq: Carga integrada de una sucesión de ADC.
14. Trigger (disparador): Nivel mínimo establecido en unidades ADC que determina el inicio de un pulso del PMT. Se almacenan los datos siguientes al trigger y unos pocos datos previos al mismo.
15. Risetime (Tiempo de Subida): Tiempo que tarda un pulso eléctrico en alcanzar el 90% de su amplitud. En términos prácticos se puede considerar el tiempo que toma en llegar a su máximo (100%) para facilitar cálculos.
16. Tubo Fotomultiplicador (PMT): Transductor lumínico capaz de detectar emisiones de luz provenientes de un fotón individual, gracias a la amplificación mediante la producción de fotoelectrones en distintas fases (generando una ganancia medible).