



## Conference Paper

# Uso de Redes Neuronales Convolucionales para el Reconocimiento Automático de Imágenes de Macroinvertebrados para el Biomonitorio Participativo

Carlos Quintero<sup>1</sup>, Fernando Merchán<sup>1</sup>, Aydeé Cornejo<sup>2</sup>,  
and Javier Sánchez-Galán<sup>1,3</sup>

<sup>1</sup>Universidad Tecnológica de Panamá, Panamá

<sup>2</sup>Grupo de Investigación en Macroinvertebrados Dulceacuícolas de Panamá. Instituto Conmemorativo Gorgas de Estudios de la Salud (ICGES)

<sup>3</sup>Instituto de Investigaciones Científicas y Servicios de Alta Tecnología AIP (INDICASAT AIP)

## Abstract

In Panama, there are community organizations that guarantee access to water for human consumption to more than 20% of the country's total population. For the sustainability of the water resource, it is essential to involve the communities in the process of monitoring the water quality. This can be achieved through the implementation of participatory biomonitoring using macroinvertebrates as indicators. In fact, it has been determined that the presence of different families of these organisms in ecosystems can be associated to different levels of their ecological quality. This work aims to develop a system capable of recognizing two families of macroinvertebrates through the use of images. The system is based on the use of algorithms of deep neural networks, with which we can achieve the learning of patterns. From a set of public images from the internet and biomonitoring carried out in the field, we train a convolutional neural network implemented in Tensorflow and Keras. These images belong to photographs of specimens of the families Calopterygidae and Heptageniidae. For this preliminary test, we report reliability percentages with values above 95%.

**Keywords:** image recognition, neural networks, convolutional neural networks, macroinvertebrates

## Resumen

En Panamá existen organizaciones comunitarias que garantizan el acceso al agua para consumo humano a más de un 20% de la población del país. Para la sostenibilidad del recurso hídrico, es indispensable darles participación a las comunidades al proceso de monitoreo de la calidad del mismo. Esto se puede lograr a través de la implementación

Corresponding Author:

Javier Sánchez-Galán  
javier.sanchezgalan@  
utp.ac.pa

Received: 15 November 2017

Accepted: 5 January 2018

Published: 4 February 2018

Publishing services provided  
by Knowledge E

© Carlos Quintero et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Selection and Peer-review  
under the responsibility of the  
ESTEC Conference Committee.



del biomonitorio participativo, empleando macroinvertebrados como indicadores. En efecto, se ha determinado que la presencia de diferentes familias de estos organismos en ecosistemas se puede asociar a diferentes niveles de calidad ecológica de los mismos. Este trabajo tiene como objetivo desarrollar un sistema capaz de reconocer dos familias de macroinvertebrados mediante el uso de imágenes. El sistema está basado en el uso de algoritmos de redes neuronales profundas (Deep Neural Networks), con el que podemos lograr el aprendizaje de patrones. A partir de un conjunto de imágenes públicas provenientes del internet y de biomonitorios realizados en campo, se aplicó el entrenamiento de una red neuronal convolucional implementado en Tensorflow y Keras. Estas imágenes pertenecen a fotografías de especímenes de las familias Calopterygidae y Heptageniidae. Para esta prueba preliminar se reportan porcentajes de confiabilidad con valores por encima del 95%.

**Palabras claves:** reconocimiento de imágenes, redes neuronales, redes neuronales convolucionales, biomonitorio, macroinvertebrados.

---

## 1. Introducción

Redes Neuronales Profundas (o Deep Networks en inglés) es el nombre que se da a las nuevas arquitecturas de las redes neurales y a los nuevos algoritmos que se usan para aprender usando estas arquitecturas. El aprendizaje profundo moderno proporciona un marco muy poderoso para el aprendizaje supervisado. Estas nuevas arquitecturas de redes han logrado cambios importantes en la dirección en la que se desarrolla la inteligencia artificial al proveer diferentes estructuras y nuevos algoritmos para el aprendizaje que permiten incrementar el número de capas y así proveyendo de mayor flexibilidad a los modelos para el reconocimiento de patrones (Goodfellow et al., 2012).

### 1.1. Uso de Redes Neuronales Para El Análisis De Imágenes

Para el reconocimiento de imágenes se hace uso de las redes neuronales convolucionales (CNNs por sus siglas en inglés), las cuales son un modelo donde las neuronas corresponden a campos receptivos de una manera muy similar a las neuronas de la corteza visual primaria de un cerebro biológico (Goodfellow et al., 2012).

La red se compone de múltiples capas. En el principio se encuentra la fase de extracción de características compuesta de neuronas convolucionales y de reducción.

A medida que se avanza en la red se disminuyen las dimensiones activando características cada vez más complejas. Al final se encuentran neuronas sencillas para realizar la clasificación.

Desde los 80's existían ideas sobre la implementación de CNNs, sin embargo, es hasta 1998 con un artículo seminal de Yann LeCun (LeCun et al., LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998), que se presentan las ideas más importantes de esta tecnología para el reconocimiento de imágenes utilizando redes neuronales convolucionales y retropropagación.

El éxito de estos sistemas se debe en gran parte a la evolución de la capacidad computacional (Cireşan et al., 2010). La inteligencia artificial actualmente permite la solución de problemas en diversos campos que requieren un alto costo computacional. Estos problemas van desde el diseño de asistentes personales, traductores, vehículos autónomos, hasta sistemas de recomendaciones de productos multimedia y compras en línea entre otros (Forbes, 2017).

Esto queda claramente demostrado en 2012 cuando gracias al desarrollo de AlexNet, una red neuronal que logró un 15% de error en la identificación de imágenes para el concurso ILSVRC (ImageNet Large Scale Visual Recognition Competition) donde su rival más cercano marcó 25% (Krizhevsky et al., 2012). El desarrollo de técnicas de aprendizaje profundo permitió que en la edición 2015 del concurso se lograra 4.8% de error, excediendo la precisión humana de 5.1% en esta misma prueba (Ioffe and Szegedy, 2015).

## 1.2. Biomonitorio participativo de la calidad del agua

La problemática de la contaminación de ríos ha llevado a los especialistas a explorar herramientas innovadoras para la evaluación de la calidad de los afluentes superficiales. Una de las alternativas que está tomando cada vez más relevancia en Panamá y otras partes del mundo, es el empleo de los macroinvertebrados como bioindicadores. Esto es debido a su amplia distribución, sedentarismo, sensibilidad a perturbaciones ambientales, largos ciclos de vida dentro del agua, gran tamaño y a que existen numerosos métodos de evaluación, tales como índices bióticos y de diversidad.

En efecto, las diferentes especies de estos organismos presentan condiciones especializadas y debido a esto representan una manera sencilla de evaluar la calidad ecológica de estas aguas (Saldaña et al. 2001). Utilizando como base la observación de estos organismos se puede realizar un estudio ambiental en comunidades remotas en contraste a alternativas más costosas como el uso de sensores. Esta metodología

de biomonitorio ofrece múltiples ventajas para la vigilancia rutinaria de la calidad del agua en las cuencas y ríos en general, ya que son simples, de bajo costo de aplicación, además de que los resultados se obtienen con gran rapidez y con una alta confiabilidad.

El problema de reconocimiento de macroinvertebrados basado en imágenes se ha estudiado anteriormente. En (Boyle et. al, 2008) se extraen diferentes características de las imágenes y se utilizan redes neuronales de una manera jerárquica y particionada. En (Tirronen et al., 2009) se implementa un algoritmo para reconocimiento de especies de macroinvertebrados. Ellos utilizan la técnica de Máquinas de Vectores de Soporte (o Support Vector Machine - SVM) que es una representación de los datos como puntos en el espacio esparcidos de una manera que facilite una clasificación binaria. En (Dietterich et al, 2009), se propone un sistema con reconocimiento de patrones en 2-D y 3-D. En este contexto, es pertinente explorar otros enfoques y otros algoritmos.

El presente trabajo tiene por objetivo desarrollar un sistema capaz de reconocer dos familias de macroinvertebrados mediante el uso de imágenes. El sistema está basado en el uso de algoritmos de redes neuronales profundas (Deep Neural Networks) y en específico el uso de Redes Convolucionales, para distinguir patrones en imágenes de macroinvertebrados.

El resto de este documento está organizado de la siguiente manera. En la Sección 2 se presenta la problemática a resolver, así como la fuente de imágenes y la metodología propuesta. En la Sección 3 se presentan los resultados obtenidos utilizando las redes de convolucionales para el análisis de las imágenes. Finalmente, en la sección 4 se presentan las conclusiones y perspectivas futuras de este trabajo.

## 2. Metodología

### 2.1. operaciones principales de redes convolucionales

Existen cuatro operaciones principales que conforman las redes convolucionales.

#### 2.1.1. Capa convolucional

Su principal propósito es extraer características de una imagen. Consiste de un conjunto de filtros entrenables que realizan producto punto con los valores de la capa precedente. En la práctica, los valores de los filtros son aprendidos para su activación

al encontrar ciertas características. Al ser colocados en cascada se obtienen diferentes niveles de abstracción.

### **2.1.2. Rectificador Lineal de Unidad**

Son utilizados después de cada convolución. Son una operación que reemplaza los valores negativos por cero y su propósito es agregar no linealidad al modelo, eliminando la relación proporcional entre la entrada y salida.

### **2.1.3. Pooling**

Algoritmo utilizado para reducir las dimensiones, con el objetivo de disminuir los tiempos de procesamiento reteniendo la información más importante.

### **2.1.4. Capa totalmente conectada (Dense Layer)**

Realiza la clasificación basado en las características extraídas por las capas de convolución y las reducidas por pooling. En esta capa todos los nodos están conectados con la capa precedente.

## **2.2. Funcionamiento De La Red Convolutiva**

Las capas de convolución y las de pooling se encargan de extraer características mientras que la capa totalmente conectada actúa como clasificador. Para el funcionamiento de este modelo debemos proceder al entrenamiento. Esto implica: 1) Inicializar todos los parámetros o pesos con valores aleatorios; 2) Utilizar una imagen de entrenamiento y utilizarla en el modelo; 3) Calcular el error total de las probabilidades resultantes del modelo y finalmente; 4) Propagar hacia atrás para calcular el error de gradiente de todos los pesos en la red y utilizar gradiente descendiente para actualizar estos valores y minimizar el error de salida.

## **2.3. Transferencia de Conocimiento entre redes convolucionales**

En la práctica es muy difícil entrenar un modelo desde cero. Esto se debe a que es difícil encontrar conjuntos de datos lo suficientemente grandes como para lograr una

buena precisión en las predicciones debido al sobreajuste que sufren las redes neuronales. Aquí es cuando debemos aplicar una técnica conocida como transferencia de conocimiento: Esta se basa en el uso de modelos previamente entrenados (Oquab et al., 2014).

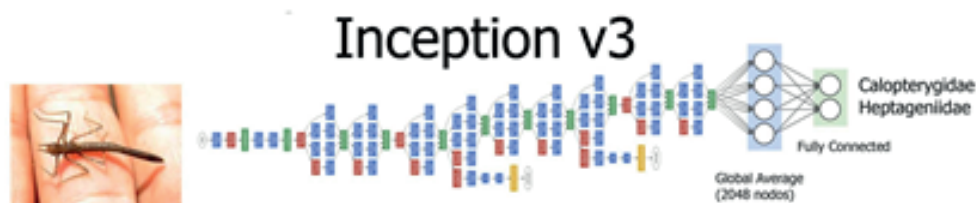
En las redes neuronales de las primeras capas obtenemos características de bajo nivel como los bordes para luego, en las capas posteriores, capturar las de alto nivel. Al utilizar modelos previamente entrenados, aprovechamos las características de bajo nivel y resolvemos el problema del sobreajuste. Además, reducimos la carga de entrenamiento que tienen un alto costo computacional para los modelos más complejos.

En este trabajo se creó una instancia de la red convolucional Inception-v3 (Szegedy et al., 2016). junto con sus valores de entrenamiento usados para la competencia ImageNet, donde se deben clasificar las imágenes en 1000 diferentes clases. Para nuestro estudio se reemplazó únicamente la capa completamente conectada con una que conecte solo a nuestras categorías y se procede a entrenar. Posteriormente se procede a un proceso de afinado donde se habilita el entrenamiento de las capas convolucionales superiores. Esto permite especializar la red neuronal para la tarea.

En Inception-v3 se utilizan capas convolucionales de tamaño 1x1 y 3x3 con su parámetro de salto en uno y dos. Adicionalmente se cuenta con una capa de pooling de valor máximo y una concatenación final de los filtros.

## 2.4. Algoritmo Inception-v3

En Inception-v3 se hace uso de unos módulos llamados Inception. Estos actúan como múltiples filtros aplicados a un mismo valor de entrada mediante capas convolucionales y de pooling. Esto permite sacar provecho de la extracción de patrones que brindan diferentes tamaños en los filtros. Luego, el resultado de estos filtros es concatenado y utilizado como el valor de salida del módulo. Este modelo aumenta el número de parámetros entrenables y la computación requerida, pero mejora considerablemente la precisión. En la Figura 1 se muestra el esquema de entrenamiento y clasificación de Inception-v3.



**Figura 1:** Esquema del modelo de clasificación Inception V3. [Adaptada de Szegedy et al., 2015]

## 2.5. Base De Datos De Imágenes

Para el entrenamiento del algoritmo Inception V3 se utilizó una base de datos de 200 imágenes. Se utilizaron imágenes obtenidas a partir de la web y fotografías obtenidas a través del monitoreo en campo.

Las imágenes de entrenamiento fueron almacenadas, para su posterior uso, en carpetas organizadas por familia y en formato JPG. Las imágenes tienen tamaños superiores a los 300 píxeles en cada una de sus dimensiones. Finalmente, todas las imágenes fueron tratadas mediante software, de manera automatizada, para normalizar los valores de los píxeles y ajustar las dimensiones de entrada definidas de acuerdo al modelo de red neuronal aplicado y agregando distorsión a las imágenes. La dimensión final que se escogió fue 224x224 píxeles.

Las imágenes fueron agrupadas de acuerdo a su especie (Calopterygidae y Heptageniidae) y divididas utilizando la proporción de 70% de ellas para entrenamiento y un 30% para validación. La selección se realizó de manera aleatoria. En la Figura 2 se muestran dos ejemplos de imágenes de la base de datos.



**Figura 2:** Muestras de la base de datos de imágenes. A la izquierda se muestra una imagen de Calopterygidae y a la derecha una de Heptageniidae.

## 3. Resultados

Luego de la estructuración de la base de datos de imágenes con carpetas separadas para cada especie, se realizó una evaluación preliminar en la tarea de clasificación utilizando el modelo Inception-v3. Para la construcción del modelo se utilizaron las librerías Tensorflow y Keras, las cuales permiten aplicar funciones matemáticas para el procesamiento de los datos y el entrenamiento del modelo. Para analizar variantes en el modelo utilizamos TensorBoard, un complemento que permite graficar los elementos que conforman el sistema y estudiar el comportamiento de los modelos utilizados.

Con la aplicación de la arquitectura Inception-v3 y la transferencia de conocimiento utilizado para ImageNet se logró clasificar las familias de macroinvertebrados

*Calopterygidae* y *Heptageniidae* con una precisión superior al 90% según el conjunto de validación. Posteriormente, mediante el proceso de afinación se logró aumentar la certeza. Esto es evidente debido a la reducción del valor de pérdida.

En la Figura 3, se muestra la gráfica de precisión vs épocas del algoritmo. Se muestra en rojo los valores obtenidos para las imágenes de entrenamiento y en verde los resultados obtenidos para las imágenes de validación. Cabe destacar que el algoritmo obtiene valores superiores al 90% en ambos casos luego de ser entrenado en 10 bloques de 32 iteraciones (épocas 0 a 9).

Luego de esto se le aplica un proceso de afinado o *fine tuning* (épocas 10 a 26) que en el caso de las imágenes de entrenamiento logra obtener mejores resultados, con un valor de 98%. Por su parte, en el caso de las imágenes de validación se llega a un máximo de 92% luego de 23 épocas, y luego baja a 89% luego de 26 épocas.

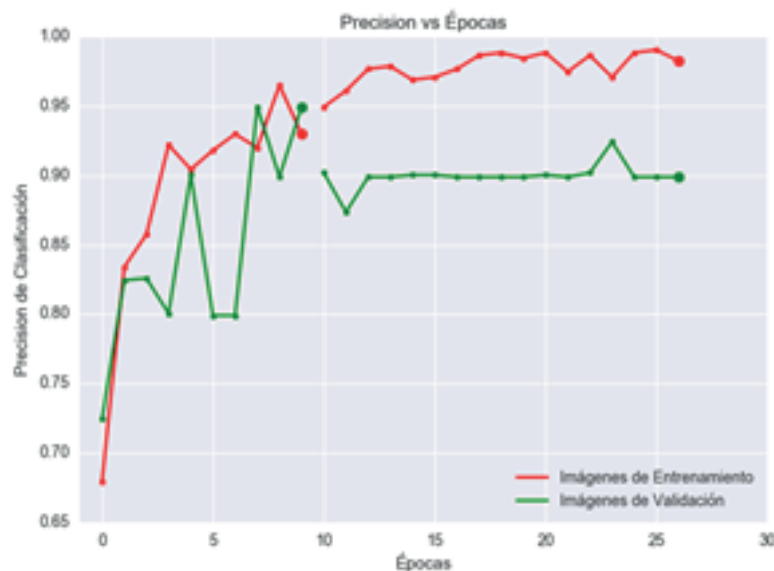


Figura 3: Precisión de Clasificación

En la Figura 4, se muestra la gráfica de error vs pasos del algoritmo. Al igual que en la Figura 2, se muestra en rojo los valores obtenidos para las imágenes de entrenamiento y en verde los resultados obtenidos para las imágenes de validación. El algoritmo obtiene valores de error de clasificación (*loss*) inferiores al 20% en ambos casos luego de ser entrenado en 10 bloques de 32 iteraciones.

Luego de proceso de afinado (*fine tuning*), en el caso de las imágenes de entrenamiento, se logra obtener mejores resultados con un valor de 6%. En el caso de las imágenes de validación, el rendimiento empeora entre otras razones a que sufre sobre-entrenamiento (*overfitting*) sobre el conjunto imágenes. También se ha sugerido que esto sucede por falta de imágenes con características novedosas para el algoritmo.





Figura 4: Error de Clasificación

Por esto, parte de nuestros objetivos futuros es recopilar una base de datos con mayor número de instancias de imágenes.

Es necesario agregar que el modelo Inception-v3 implementado en su totalidad ocupa 87.23MB de espacio por iteración. Esto corresponde a los 21,806,882 parámetros flotantes de 32 bits que para el entrenamiento es utilizado en bloques de 32 sumando un total de 2981.28MB de memoria. Además, para el modelo Inception-v3 utilizado se calculan  $2.35E+09$  FLOP en las capas convolucionales. Sin embargo, en el proceso de entrenamiento sólo se consideran las últimas 279 capas lo que corresponde a un costo computacional de  $7.77E+08$  FLOP.

## 4. Conclusiones

Al construir un modelo basado en Inception-v3 entrenado para ILSVRC-2012 se demostró la posibilidad de aplicar el proceso de transferencia de conocimiento para alcanzar resultados acordes con el estado del arte sin necesidad de un conjunto de datos grande o de equipos especializados de computación. Los resultados presentados son arriba del 90% para el conjunto de imágenes de entrenamiento y validación. Por su parte el error de clasificación se llevó hasta el 6% en el conjunto de imágenes de entrenamiento, pero sucumbió al sobre-entrenamiento.

En la gráfica para error de clasificación se aprecia una reducción por debajo de 0.1 para las imágenes de entrenamiento sin recibir un incremento notable en la precisión

de validación. Esto es un indicativo de que el modelo está sufriendo de sobreajuste. Para mejorar dicho problema se requiere de un incremento en la base de datos de imágenes. Esta base de datos puede ser alimentada a medida que se implementa el proyecto de biomonitorio participativo, brindando soporte para una mejora continua del sistema.

Podemos agregar que, con la aplicación de los algoritmos de inteligencia artificial más recientes, que permiten la transferencia del entrenamiento y el entrenamiento solamente en las últimas capas, se prevé que en futuras versiones de este trabajo se incluya una variedad más amplia de familias de macroinvertebrados. Por ejemplo, no solo utilizar las familias Calopterygidae y Heptageniidae, sino clasificar 81 familias que están presentes en nuestro país y así proveer una herramienta poderosa y que pueda llevar a cumplir los objetivos del proyecto de biomonitorio participativo con la comunidad.

## 5. Agradecimientos

Este trabajo es parcialmente financiado por la Secretaría Nacional de Ciencia, Tecnología e Innovación de Panamá (SENACYT) a través del Proyecto IDDS15-054, titulado Biomonitorio Participativo de la Calidad del Agua con Juntas Administradoras de Acueductos Rurales ([gorgas.gob.pa/aplicaciones/biomonitorio/](http://gorgas.gob.pa/aplicaciones/biomonitorio/)). Además, la investigación es parcialmente financiada por el Sistema Nacional de Investigación (SNI) de SENACYT, del cual el co-autor F. Merchán es miembro (contrato No.124-2015).

## Referencias

- [1] Brownlee, J. (2016). 8 Inspirational Applications of Deep Learning.
- [2] Boyle, S., Somma, A., Codjia, J., Ure, J., Romanelli L., & Momo, F. (2008). Recognition of freshwater macroinvertebrate taxa by image analysis and artificial neural networks, In Proceedings of the International Symposium on Mathematical and Computational Biology, Campos do Jordão, Brazil, 22-27 November 2008.
- [3] Cireşan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. *Neural computation*, 22(12), 3207-3220.2
- [4] Coutinho, E., Weninger, F., Schuller, B. W., & Scherer, K. R. (2014, October). The Munich LSTM-RNN Approach to the MediaEval 2014 "Emotion in Music" Task. In MediaEval.

- [5] Dietterich, T.G., Lytle, D.A., Mortensen, E.N., Moldenke, A.R., Paasch, R.K., Sarpola, M.J., & Shapiro, L.G. (2009). An Aquatic Insect Imaging System to Automate Insect Classification. Transactions of the American Society of Agricultural and Biological Engineers.
- [6] Forbes. (2017, January). 10 Powerful Examples of Artificial Intelligence in Use Today. (R. Adams, Productor).
- [7] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning (pp. 448-456).
- [8] Goodfellow, Ian, Bengio, Yoshua and Courville, Aaron. (2016, December). Deep Learning. MIT Press.
- [9] Karpathy, A. (2015). The unreasonable effectiveness of recurrent neural networks. Andrej Karpathy blog.
- [10] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
- [11] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
- [12] Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1717-1724).
- [13] Saldaña Fabela, M. d., López López, R., & Salcedo Sánchez, E. (2001). *Utilización de un índice de diversidad para determinar la calidad del agua en sistemas lóticos*. Ingeniería hidráulica en México, Abr-jun, pp. 57-66, 2001.
- [14] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D.,... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-9).
- [15] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2818-2826).
- [16] Tirronen, V., Caponio, A., Haanpää, T., & Meissner, K. (2009). *Multiple Order Gradient Feature for Macro-Invertebrate Identification Using Support Vector Machines*. Adaptive and Natural Computing Algorithms. ICANNGA 2009. Lecture Notes in Computer Science, vol 5495. Springer, Berlin, Heidelberg.



## Authorization and Disclaimer

*Authors authorize ESTEC to publish the paper in the conference proceedings. Neither ESTEC nor the editors are responsible either for the content or for the implications of what is expressed in the paper.*