

## Consolidating INIA's Rice Breeding Program Database, Phase I: Historical Indica Trials

S Scheffel<sup>1</sup>, J Rosas<sup>2</sup> and F Pérez<sup>3</sup>

<sup>1</sup> Maestría en Ciencias Agrarias, Facultad de Agronomía, Universidad de la República, Uruguay  
scheffelsheila@gmail.com

<sup>2</sup> Instituto Nacional de Investigación Agropecuaria (INIA), Uruguay, jrosas@inia.org.uy

<sup>3</sup> Instituto Nacional de Investigación Agropecuaria (INIA), Uruguay, fperez@inia.org.uy

### ABSTRACT

INIA's Rice Breeding Program (IRBP) is the major rice breeding program in Uruguay. To deliver improved cultivars, the performance of thousands of experimental lines is assessed in field and laboratory trials. The joint analysis of all the experimental data, combining multiple years, locations and trials, allows better estimations of the genetic value of the breeding lines and thus optimizes the efficiency of the breeding program. Records of historic field and laboratory trials of the IRBP are fragmented in a multiplicity of media and formats that prevent their joint analysis. This work describes the integration of a database corresponding to the Indica subtype materials, which constitutes approximately half of the IRBP germplasm. It comprises data from field and lab trials across 14 years (2006 to 2019) and three locations. The strategy used was: 1) Implementation of uniform templates with standardized field names. 2) Manual review and quality control of data (variable identification and assignation to a standardized field, check units of measurement, check cell format, identification of missing data, and standardized annotation of qualitative observations). 3) Reading of the formatted spreadsheets of each trial with the R software, consolidating all trials in a single data frame with all the Indica IRBP data. 4) Standardization of field values (missing data and standardized levels for categorical variables). 5) Trial-wise quality control with statistical criteria (distribution of variables, trial heritability, spatial distribution of residuals). A database with around 26,000 records was obtained, comprising about 200 trials. Consolidated data will be crucial for implementing multiple-environment analysis that will enhance selection accuracy and improve genetic gain in the IRBP.