www.journalworlds.com
AGRI JOURNAL WORLD
VOLUME 4 ISSUE 1 JANUARY, 2024

R SOFTWARE: AN EFFECTIVE OPEN-SOURCE TOOL FOR ADDRESSING INTEGRATED WATER MANAGEMENT IN AGRICULTURE

Naveena K*, Santosh Onte, Sharannya T M and Ashish K. Chaturvedi

Centre for Water Resource Development and Management

*Corresponding author email: naveenak@cwrdm.org

ABSTRACT

Managing agricultural water resources in the context of global climate change is a complex challenge. A comprehensive understanding of hydrological processes is crucial for sustainable water management. Hydrological models are pivotal in assessing water availability, runoff, and groundwater recharge, and estimating evapotranspiration (ET) accurately. Integrating plant physiological tools with ET models allows precise crop water requirement estimation, while crop models simulate the impact on crop growth and water usage. Despite challenges in tool integration, R programming, as an open-source platform, offers a transformative solution. This article explores R's potential for predictive analytics, enabling precise water scheduling, irrigation optimization, crop forecasting, and climate impact assessment, making it an ideal platform for efficient and resilient water management.

KEYWORDS Agricultural Water Management, Climate Change, Decision Support System, Hydrological Process, R Programming

INTRODUCTION

Managing water in agriculture is crucial for dealing with global water shortages and securing enough food sustainably. Understanding how crops use water at different stages is key to using water smartly in agriculture. It can achieve by adopting improved irrigation methods and accurately assessing crop water requirements. India, having 18% of the world's population but possessing only 4% of its water in world scenario (Kumar, 2019) leads to water scarcity in few regions. This scarcity is intensified by erratic monsoons and climate change impacts, leading to significant water shortages for a large portion of the population (Sharannya et al., 2021). Groundwater depletion due to excessive agricultural use further compounds this issue. Extensive research in India is imperative to develop efficient irrigation techniques, enhance water productivity, predict rainfall accurately, and manage surface and groundwater effectively. Effective management of agricultural water resources demands a comprehensive understanding of hydrological processes using hydrological models like SWAT to assess water availability, runoff, and groundwater recharge. Precise estimation of evapotranspiration (ET) plays a pivotal role in effective water management strategies. By merging plant physiological tools with robust ET models like the Penman-Monteith equation, we gain a refined approach to calculating crop water needs. Similarly, employing crop models like DSSAT, CROPWAT helps simulate the impact of diverse factors on crop growth and water utilization. This integrated approach ensures not only crop-specific irrigation practices but also considers the broader water balance of the region. However, integrating these tools into a cohesive decision support system requires substantial effort and learning for researchers.

R programming, an open-source platform, emerges as a transformative solution to these challenges. This article aims to explore how researchers and stakeholders can leverage the multifaceted toolkit of R. It empowers predictive analytics, precise water scheduling, optimized irrigation strategies, crop forecasting, and climate impact assessments. Beyond analytics, R facilitates user-friendly decision support systems, guiding stakeholders through critical choices in water management strategies. The adoption of R becomes a strategic necessity due to recent advancements in R packages addressing water security and climate complexities, making it an ideal platform for efficient estimation of results and nurturing resilience in water management. The figure 1 illustrates the key focus areas in agricultural water management. The workflow in Figure 2 helps researchers and students pick the suitable analyses and R package for making effective water management strategies.

VISUALIZING AND HANDLING OF AGRICULTURAL WATER DATA WITH R'S DIVERSE PACKAGES

Data processing and visualizing holds immense significance for efficient analysis and interpretation. Data preparation to make it suitable for further treatment is very necessary to reduce the input uncertainties. The package like "**dplyr**" helps for data processing and manipulation such that further analysis can be easily done by the researchers. We can easily select, mutate, and replace the required data and variable using this package (Fischetti, 2018).

The creation of diverse indices for meteorological parameters such as rainfall, temperature, and evapotranspiration is pivotal in efficient water management strategies. The "**ClimPACT2**" package offers an R-based dashboard framework specifically designed for analyzing rainfall and temperature indices (Alexander & Herold, 2016). This tool facilitates a comprehensive assessment of these meteorological factors, enabling insights crucial for water resource management decisions. The "**Evapotranspiration**" package is a powerful tool that uses 17 models to estimate actual, potential, and reference crop evapotranspiration (Guo et al, 2016). This comprehensive resource is essential for understanding and managing water needs in crops and ecosystems.

Figure 1: Plausible areas to address efficient Agricultural Water Management

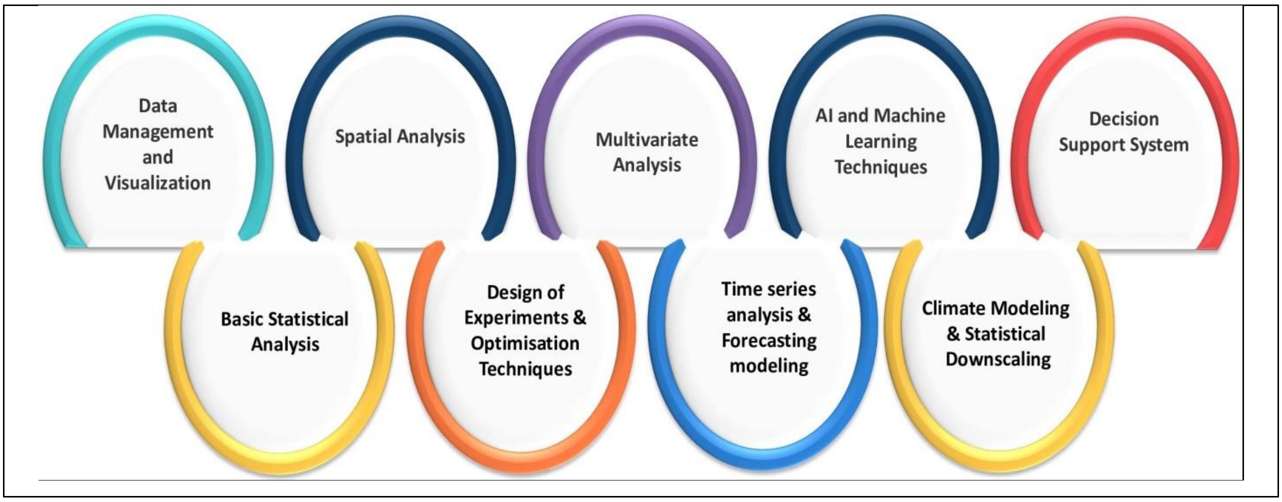

Figure 2: **Strategic Flow Utilizing R Packages for Enhanced Assessment of Agricultural Water Management**

Further visual representations facilitate effective communication of information among stakeholders about irrigation scheduling, resource allocation, drought management strategies etc. Visualizations enable the identification of long-term trends, empowering proactive planning for changing climatic conditions. **'ggplot2'** meant for static visualization of almost all type of datasets (*Wickham et al, 2016*). **'plotly':** Delivers interactive geographical insights, allowing users to explore spatial patterns and data correlations interactively (*Sievert 2020*). **'leaflet':** Specializes in dynamic maps, facilitating the visualization of spatial information so we can clearly visualize the water resources, soil moisture, and crop distribution (*Graul, and Graul 2016*). **'rasterVis':** Aids in monitoring crop health using satellite imagery, enabling the visualization and analysis of raster data (*Lamigueiro et al, 2023*).**'highcharter':** Offers interactive charts, aiding in the exploration of complex data relationships and patterns (*Smith et al, 2022*).

BASIC STATISTICAL TOOLS FOR AGRICULTURAL WATER MANAGEMENT IN R

Basic statistical analyses like descriptive statistics, statistical inference, correlation, and regression are vital for understanding relationships, making predictions, and informing decision-making.

DESCRIPTIVE STATISTICS: The packages like '**summarytools**' (*Comtois and Comtois, 2016*), '**psych**' **(**Revelle and Revelle, 2015) Offers functions to generate descriptive statistics techniques like mean, median, standard deviation, variance, etc. it helps for summarization of data connected to water usage, rainfall, crop yields, and other relevant variables.

STATISTICAL INFERENCE: '**stats**', '**infer**' packages facilitate hypothesis testing, estimating parameters, and constructing confidence intervals (*Couch et al, 2021*). They help assess significant difference in factors affecting water productivity or impact of irrigation methods on crop yield.

CORRELATION AND REGRESSION ANALYSIS: **'corrplot' (*Wei et al, 2017*) 'Mass' (*Ripley et al 2013*)** Packages enable to calculate and visualization of correlations between variables, such as the relationship between rainfall and crop yield or water usage and soil moisture. Spearman rank correlation and Pearson correlation can be assessed by this package in effective manner. The '**lm**' function performs linear regression, for building functional relationship between response and predictor variables. '**Mass**' aids in developing stepwise regression model especially during the issue of multi-collinearity.

DESIGN OF EXPERIMENTS (DOE) AND OPTIMIZATION TECHNIQUES

Design of Experiments (DoE) systematically plans and conducts trials in agricultural water management. Both block and non-block design can be adopted to evaluate the factors under consideration. Optimization techniques, like linear programming, genetic algorithms, and response surface methodology determine optimal irrigation schedules, crop water needs, and allocation strategies, ensuring maximal water productivity by considering factors such as soil types, weather forecasts, and crop requirements. By integrating DoE and optimization, practitioners enhance water-related variable exploration, leading to improved resource efficiency, better crop yields, and sustainable water practices.

a. *DESIGN OF EXPERIMENTS (DOE):* '**DoE.base**' provides tools for creating experimental designs, aiding in planning efficient trials for studying water-related variables (*Gromping et al, 2018*). '**agricolae**' and '**multicompView**' facilitates the implementation of various experimental designs in agricultural studies, including irrigation and crop-related experiments (*Mendiburu et al,2015*).

b. *OPTIMIZATION TECHNIQUES:* '**GA**' is specifically for genetic algorithms (*Scrucca, 2013*), '**lpSolve**' for linear programming (*Musunuru, 2014*), and '**NMOF**' provides various numerical methods for optimization (*Schumann et al, 2013*). '**rsm**' Specifically designed for Response Surface Methodology, this package helps in designing experiments, fitting response surface models, and conducting analyses to optimize processes (*Lenth, 2015*).

TIME SERIES ANALYSIS IN AGRICULTURAL WATER MANAGEMENT

Time series analysis in agricultural water management involves studying data collected over time to understand patterns, trends, and fluctuations in water-related variables. Time series analysis predicts future water availability, identifies seasonal trends, and optimizes irrigation schedules, crucial for efficient water usage in agriculture. It helps understand historical patterns, aiding in informed decision-making and adaptation to changing environmental conditions for sustainable water management (Naveena et al,2023). R packages aid in time series analysis for agricultural water management: **'ts', 'forecast', 'TSA', 'imputeTS','trend' 'xts'** and '**zoo**'

The **'ts'** is basic package held for handling time series data in R, it enable simple time series analysis and visualization. **'imputeTS'** package helps for imputing missing values in the time series (Kannegowda et al, 2023). **'TSA'**: Provides functions for time series analysis, including decomposition, smoothing, and modeling of seasonal components, aiding in understanding seasonal patterns in series. The '**trend**' and '**modifiedmk**' package used for identifying long term pattern specially for weather parameter (Naveena et al, 2023; Tasiya et al, 2023; Elzopy et al, 2021). Where **'forecast'** package Offers models for forecasting future, this is vital for predicting water availability, crop water requirements, rainfall patterns etc (Naveena et al, 2017). **'xts'** and '**zoo**': These packages handle irregular time series data efficiently, and useful to deal missing observation.

STATISTICAL DOWNSCALING

Statistical downscaling translates broad climate model data to local-scale projections crucial for water management decisions. Techniques like Quantile Mapping, Regression- and Artificial Neural Networks based models establish relationships between large-scale climate variables and local-scale data, aiding in precise projections. These methods adjust model data to match historical observations, empowering hydrologists and water managers to make informed decisions regarding water resources, flood control, and agricultural water management at a local level. There are several R packages that facilitate statistical downscaling methods beneficial for agricultural water management (Paz, & Willems 2022). The **'downscale':** Offers functions to perform statistical downscaling, particularly for climate data, allowing users to translate large-scale climate variables to finer local scales.**'qmap':** Specifically designed for Quantile Mapping, a statistical downscaling method that adjusts model data distribution to match observed data's quantiles, often used for variables like precipitation and **'downscaleR':** Provides tools for climate data downscaling, focusing on generating local-scale projections from large-scale climate models.

SPATIAL ANALYSIS IN AGRICULTURAL WATER MANAGEMENT

Spatial analysis in agricultural water management involves studying and analyzing spatially distributed data connecting to water resources, soil moisture, crop distribution, and more. It helps in understanding patterns, relationships, and variations across geographical areas, aiding in decision-making for water management. The '**sp**' is the fundamental package for handling spatial data, allowing for spatial data classes, visualization, and basic spatial operations. '**raster**' is used to deal raster data (gridded data), beneficial when dealing with spatial information such as satellite imagery, land use, or soil properties. '**sf**' : Modern package for handling spatial data in R, particularly for vector data (points, lines, polygons), enabling spatial operations and visualization. '**gstat**': Facilitates geo-statistical analysis, including spatial interpolation methods that estimate values at unsampled locations based on nearby observations, valuable for predicting soil moisture or groundwater levels. '**spdep**' : Focuses on spatial dependence and spatial regression, aiding in exploring relationships between spatially referenced variables, helpful when studying the impact of geographical factors on water resources or crop yield.

MULTIVARIATE ANALYSIS FOR AGRICULTURAL WATER MANAGEMENT

In the context of agricultural water management, multivariate analysis can help in examining complex relationships between various factors affecting water resources, crop growth, and irrigation strategies. R, packages and methods for conducting multivariate statistical analysis: **'FactoMineR', 'stats', 'Caret', and 'MASS'.** The **'stats':** Offers foundational functions for multivariate analysis, including methods for principal component analysis (PCA), factor analysis, and clustering. **'FactoMineR':** Provides tools for multivariate exploratory data analysis, PCA, factor analysis, and correspondence analysis, valuable for understanding complex relationships between multiple variables in agricultural datasets (*Joseph et al, 2023*). **'Caret':** While mainly used for machine learning, **caret** includes functions for pre-processing data, feature selection, and building predictive models across multiple variables. **‘MASS’:** Offers various multivariate analysis functions, including linear and quadratic discriminant analysis, useful for classification problems in agricultural studies.

AI AND MACHINE LEARNING FOR AGRICULTURAL WATER MANAGEMENT

AI and machine learning revolutionize agricultural water management, predicting crop yield and optimizing irrigation through regression and genetic algorithms. Techniques like LSTM networks forecast water needs, while image processing monitors crop health via satellite data. Decision trees, Support vector machine, ANN and Random forest uncover factors influencing water consumption, drought assessment, assessing impacting crop health for better management (Sharannya et al., 2022). These tools collectively enhance precision, efficiency, and sustainability in agricultural water management. R packages specifically tailored for AI and machine learning applications (*Lantz et al,2019; Tyagi, & Chahal, 2022*) **‘caret’:** Provides a unified interface for various machine learning algorithms and facilitates model training, testing, and tuning. **‘randomForest’:** Implements Random Forest algorithms for classification and regression tasks. **‘xgboost’ and ‘gbm’:** Implement gradient boosting algorithms, aiding in predictive modeling and regression tasks, offering improved performance. **‘rpart’ and ‘rpart.plot’:** Facilitate decision tree algorithms, aiding in understanding factors influencing water consumption and resource allocation. **‘keras’ and ‘tensorflow’:** Enables the implementation of deep learning models, including neural networks like LSTM networks, beneficial for time series forecasting such as predicting water needs. **‘glmnet’:** Assists in implementing Lasso and Ridge regression, useful for feature selection and regularization in predictive modeling. **‘e1071’:** Provides tools for support vector machines (SVM), beneficial for classification and regression tasks in water management datasets. **‘H2O’:** Offers machine learning algorithms and tools for scalable and distributed computing, beneficial for handling large datasets.

DEVELOPING DECISION SUPPORT SYSTEMS (DSS) THROUGH R

Decision Support Systems integrate diverse data sources and models, aiding decisions by simulating scenarios and offering recommendations for better water management in agriculture. These systems compile visualization of complex data, easy comprehension of statistical analysis results, and facilitate real-time monitoring for adaptive strategies. Users interact with interfaces, adjusting parameters for optimal irrigation, crop selection etc. They assess risks and support quick responses to changing conditions, ensuring sustainable water use (*Mohamed et al, 2021*). For Agricultural Water Management, Decision Support Systems using R involve compiling several R packages capable of visualization, statistical and machine learning analyses, optimization tools, etc. These packages enable the creation of interactive web applications through tools such as '**shiny**' and associated packages like '**shinydashboard**' (for creating dashboards with various components like charts, tables, and maps) and '**shinyWidgets**' (offering additional interactive components, enhancing the user experience within the DSS interface). These packages, when used together, enable the development of interactive, user-friendly Decision Support Systems tailored for Agricultural Water Management, integrating data visualization, mapping, machine learning, and spatial analysis functionalities (*Kukar et al, 2019*).

CONCLUSION

In addressing agricultural water challenges, R programming stands as a transformative force, enabling to understand hydrological process more effectively. Development of rainfall, temperature and evapotranspiration indices, visualization, statistical analysis, advanced statistical, machine learning, optimization techniques, spatial analysis, and hydrological modelling can easily dealt using r for precise water scheduling, crop prediction, and climate impact assessment. R's also extends beyond analytics, enabling user-friendly decision support systems that guide stakeholders through sophisticated choices.

REFERENCES

Alexander, L., & Herold, N. (2016). ClimPACT2: Indices and software.

Biradar, B., Kulkarni, S. A., Shobharani, M., Sidramappa, G. B., Rathod, S., Naveena, K., ... & Arabhanvi, F. (2023). Enhancing legume crop protection: Machine learning approach for accurate prediction of lepidopteran pest populations in Kalyan Karnataka.

Comtois, D., & Comtois, M. D. (2016). Package ‘summarytools’.

Couch, S. P., Bray, A. P., Ismay, C., Chasnovski, E., Baumer, B. S., & Çetinkaya-Rundel, M. (2021). infer: An R package for tidyverse-friendly statistical inference. *Journal of Open Source Software*, 6(65), 3661.

De Mendiburu, F., & Simon, R. (2015). Agricolae-Ten years of an open source statistical tool for experiments in breeding, agriculture and biology (No. e1748). *PeerJ PrePrints*.

Elzopy, K. A., Chaturvedi, A. K., Chandran, K. M., Gopinath, G., K, N., & Surendran, U. (2021). Trend analysis of long-term rainfall and temperature data for Ethiopia. *South African Geographical Journal*, 103(3), 381-394.

Fischetti, A. (2018). Data analysis with R: A comprehensive guide to manipulating, analyzing, and visualizing data in R. Packt Publishing Ltd.

Graul, C., & Graul, M. C. (2016). Package 'leafletR'. Interactive Web-Maps Based on the Leaflet JavaScript Library. Available online: https://github. com/chgrl/leafletR (accessed on 1 April 2016).

Grömping, U. (2018). R package DoE. base for factorial experiments. *Journal of Statistical Software*, 85, 1-41.

Guo, D., Westra, S., & Maier, H. R. (2016). An R package for modelling actual, potential and reference evapotranspiration. *Environmental Modelling & Software*, *78*, 216-224.

Joseph, C. J., Renjith, K. R., Santhosh, R., Ihjas, K., & Naveena, K. (2023). Distribution of geochemical forms and bioavailability of phosphorus in the surface sediments of Beypore Estuary, southwestern coast of India. *Environmental Monitoring and Assessment*, 195(12), 1458.

Kannegowda, N., Udayar Pillai, S., Kommireddi, C. V. N. K., & Fousiya. (2023). Comparative assessment of univariate and multivariate imputation models for varying lengths of missing rainfall data in a humid tropical region: a case study of Kozhikode, Kerala, India. Acta Geophysica, 1-16.

Kukar, M., Vračar, P., Košir, D., Pevec, D., & Bosnić, Z. (2019). AgroDSS: A decision support system for agriculture and farming. *Computers and Electronics in Agriculture*, 161, 260-271.

Kumar, R. (2019). Emerging challenges of water scarcity in India: the way ahead. *International Journal of Innovative Studies in Sociology and Humanities*, 4(4), 6-28.

Lamigueiro, O. P., Hijmans, R., & Lamigueiro, M. O. P. (2023). Package 'rasterVis'.

Lantz, B. (2019). Machine learning with R: expert techniques for predictive modeling. Packt publishing ltd.

Lenth, R. V., Lenth, M. R. V., & Vdgraph, S. (2015). Package 'rsm'.

Mohamed, E. S., Belal, A. A., Abd-Elmabod, S. K., El-Shirbeny, M. A., Gad, A., & Zahran, M. B. (2021). Smart farming for improving agricultural management. *The Egyptian Journal of Remote Sensing and Space Science*, 24(3), 971-981.

Musunuru, K. (2014). Understanding Systems of Linear Equations and Programming through lpSolve and R Language.

Naveena, K., & Subedar, S. (2017). Hybrid time series modelling for forecasting the price of washed coffee (Arabica Plantation Coffee) in India. *International Journal of Agriculture Sciences*, ISSN, 0975-3710.

Naveena, K., Tasiya, R., & Rana, S. (2023). Spatio-temporal Trend Analysis of Rainfall using R Software and ArcGIS: A Case Study of an Agro-climatic Zone-1 of Gujarat, India. Springer Nature.

Paz, S. M., & Willems, P. (2022). Uncovering the strengths and weaknesses of an ensemble of quantile mapping methods for downscaling precipitation change in Southern Africa. *Journal of Hydrology: Regional Studies*, 41, 101104.

Pimentel, D., Berger, B., Filiberto, D., Newton, M., Wolfe, B., Karabinakis, E., ... & Nandagopal, S. (2004). Water resources: agricultural and environmental issues. *BioScience*, 54(10), 909-918.

Plant, R. E. (2018). Spatial data analysis in ecology and agriculture using R. cRc Press.

Revelle, W., & Revelle, M. W. (2015). Package 'psych'. The comprehensive R archive network, 337(338).

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. Cran r, 538, 113-120.

Schumann, E. (2023). Package 'NMOF'.

Scrucca, L. (2013). GA: A package for genetic algorithms in R. *Journal of Statistical Software*, 53, 1-37.

Sharannya, T. M., Venkatesh, K., Mudbhatkal, A., Dineshkumar, M., & Mahesha, A. (2021). Effects of land use and climate change on water scarcity in rivers of the Western Ghats of India. *Environmental Monitoring and Assessment*, 193(12). https://doi.org/10.1007/s10661-021-09598-7

Sharannya, T. M., Kolluru, V., Amai, M., & Acharya, T. D. (2022). Enhanced streamflow simulations using nudging based optimization coupled with data-driven and hydrological models. *Journal of Hydrology: Regional Studies*, 43. https://doi.org/10.1016/j.ejrh.2022.101190

Sievert, C. (2020). Interactive web-based data visualization with R, plotly, and shiny. CRC Press.

Smith, J., Sun, Y., Hijano, D. R., Hoffman, J. M., Hakim, H., Webby, R. J., ... & Tang, L. (2022). covidscreen: a web app and R Package for assessing asymptomatic COVID-19 testing strategies. *BMC Public Health*, 22(1), 1-11.

Tasiya, R. F., Rana, S. C., & Naveena, K. (2023). Change point and trend analysis of rainfall for the semi-arid zone of Gujarat state. *Water and Energy International*, 65(11), 6-14.

Tyagi, A. K., & Chahal, P. (2022). Artificial intelligence and machine learning algorithms. In Research Anthology on Machine Learning Techniques, Methods, and Applications (pp. 421-446). IGI Global.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). Package 'corrplot'. Statistician, 56(316), e24.

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. Create elegant data visualizations using the grammar of graphics. Version, 2(1), 1-189.

How to cite:
Naveena, K., Onte, S., Sharannya, T., and Chaturvedi, A. (2024). R software: An effective open-source tool for addressing integrated water management in agriculture. Leaves and Dew Publication, New Delhi 110059. *Agri Journal World., 4(1):7-18.*

*********XXXXX**********