



AN ANALYTICAL APPROACH TO GIS ANALYSIS AND MAP MAKING USING OPEN SOURCE SOFTWARE

Tarick Hosein¹, Bheshem Ramlal^{2*}, Lisa Kirton-Reed³ and Adrian Trotman⁴

^{1,2}Faculty of Engineering, The University of the West Indies, Trinidad

^{3,4}Caribbean Institute for Meteorology and Hydrology, Barbados

¹Email: tarick.hosein@sta.uwi.edu

²Email: bheshem.ramlal@sta.uwi.edu

³Email: lkreed@cimh.edu.bb

⁴Email: atrotman@cimh.edu.bb

Abstract: Over the last two decades, advancements in analytical cartography have reduced the need for professional geomatics experts when creating maps. GIS software have greatly simplified the map creation process, allowing non-experts to make maps using many built-in tools. However, this has resulted in a lack of quality control and assessment when data is manipulated and processed towards developing maps. This also raises questions with regards to reliability, validity and objectivity of data, results and maps. In contrast, this paper presents a methodology for automating both the GIS analysis and Map Making processes using analytical cartography in a non-standard GIS environment. The work was done primarily in the R environment, open source software and programming, to replicate processes that were normally done using expensive proprietary software and many person hours. Applied in a case study to improve the workflow efficiency in precipitation index modelling, the developed system was found to improve on overall time by tenfold, the formats and number of products generated were more than doubled, the replicability of map layouts was greatly improved, and the published formats increased.

Keywords: *Cartography, GIS, Automation, R*

<https://doi.org/10.47412/JKKP1396>

1. Introduction

Maps have always been a unique mechanism for communicating spatial information. Traditionally, cartographers have taken information and published a visualization in a form that rapidly and concisely transmits the complex spatial relationships of the 3D world onto a flat paper or computer screen [1, 2]. As such, cartography was considered to be the “art, science, and technology of making maps” [3]. The ability of cartographers to distort reality in the form of geometry (projection, scale and generalization) and information suppression also meant that maps are more than a mechanism for displaying information but capable of telling stories and manipulating the truth [2]. Monmonier [2] considered maps as intellectual tools and in some cases, weapons used to “legitimize territorial conquests, economic exploitation and cultural imperialism” especially



when used for propaganda. Notwithstanding wilful misrepresentation as part of political and other agendas, cartographers have operated within best practice systems of laws, principles, policies, lessons and recommendations as ethical professionals of data representation and presentation [1]. Cartographic rules have guided the professionals into selecting the best projection, scales, symbols, generalization, simplification, and weighting and subsequent omission of data as they concisely aggregate reality into their uncluttered flat representation. This meant that the cartographer needed to have a clear understanding of the rules of map making, the purpose of the map, and the intended audience.

Within the power of modern Geographic Information Systems (GIS), anyone can make maps. Digital cartography has become an intuitive process without regards to basic map rules and map users [1]. The “art” and “science” aspects of cartography have been replaced by technology. Technology which has further evolved into special forms of Artificial Intelligence (AI) that integrate Expert Systems (ES) of cartography within the digital software, as a component of analytical cartography. This has further assisted non-cartographers in map making by proposing appropriate solutions while using cartographic rules in deciding spatial visualizations [1]. However, such solutions are primarily available within commercial software integrating “science” and “technology” but incapable of addressing the adept “artistic” nature of cartographic decision-making and application. In contrast to commercial solutions, low cost and free expert solutions can also be developed from previous models and outcomes [1]. In essence, this means that the professional solution applied to one project, can be transferred and reapplied in another similar project.

Analytical cartography was briefly introduced in the previous section as a form of AI. However, analytical cartography is much more than a system of assisting with cartographic decision-making. Moellering [4] categorizes analytical cartography into five main groups: Maps and Coordinate Transformation (Real and Virtual); Spatial Data Models and Structures; and Spatial Frequencies, Sampling Theorem, Neighbourhood Operators, and Fractals; Map Generalization Theory; and Additional Developments in Spatial Theory. These groups have become the underlying theoretical base for many Geographic Information Systems (GIS) software, becoming the backbone of digital cartography, with some commercial software offering much more ES functionality within each group than free and open source software.

The use of open source software to facilitate customized solutions enables better tailoring of workflows processes [5]. In this way, customized solutions will ensure that workflows match the skills available while increasing productivity by improving the use of available resources and reducing dependency on external assistance [6]. This is especially the case within organisations that have limited human and financial resources. The use of open source also ensures that users are not dependent on software licenses that can lead to significant downtime between renewals and updates [5]. It is also beneficial to research oriented agencies, as there is usually a willing community of expertise to assist in development. Finally, the use of programmable solutions can significantly improve the efficiency of doing repetitive tasks while freeing up resources and people to pursue other activities.



2. Monthly Climate Monitoring Maps: A Case Study

The case study reviewed how Standardized Precipitation Index (SPI) maps were processed and created at the Caribbean Institute of Hydrology and Meteorology (CIMH) and found the following stages:

- Data Preparation – sourcing and preparing internet based satellite information representing the precipitation data as grid information. High temporal reanalysis daily data is aggregated into monthly totals, and stored in a localized “historic database.” Concurrently, a database of land-based observations precipitation totals is maintained and updated monthly.
- Analytical processing – using a customized algorithm, point information from the databases (reanalysis and land based observations) are processed to determine trends in the precipitation in the form of SPIs. The preparation of data and execution of the algorithms were executed manually for each point. This meant that the operator needed to create over 250 ASCII files using MS Excel and individually process each one in a customized statistical algorithm to extract SPI values. Each of the ASCII files contained over 18 years of monthly rainfall data, reflecting over 50K data entries.
- The outcome of the processed information resulted in a vector of values for each position representing temporal trends. For each spatial point position, the algorithm extracted SPI trends for 1 month, 3 months, 6 months, 12 months and 24 months periods. All SPIs, for the same analysis period, were aggregated into five geospatial vectors containing all the land based and reanalysis data. The spatial extent of the analysis consists of the Caribbean Sea and adjacent countries/islands.
- Each listing was then imported into the commercial software ArcGIS. Points that were very close to each other were manually removed, with priority given to the land observations, before executing a spline surface interpolation.
- Subsequently, the interpolated surface was masked to exclude some areas, categorized and coloured accordingly.
- A country boundary shapefile dataset was overlaid onto the interpolated/masked surface along with general layout information, including title, metadata, legend, and grids.
- Finally, the maps were exported as geotiff, pdfs, and other image formats for use in web and printed publishing.

The total processing time of this workflow model approximated 2.5 days. A cartographic model depicting the workflow is shown in Fig. 1.

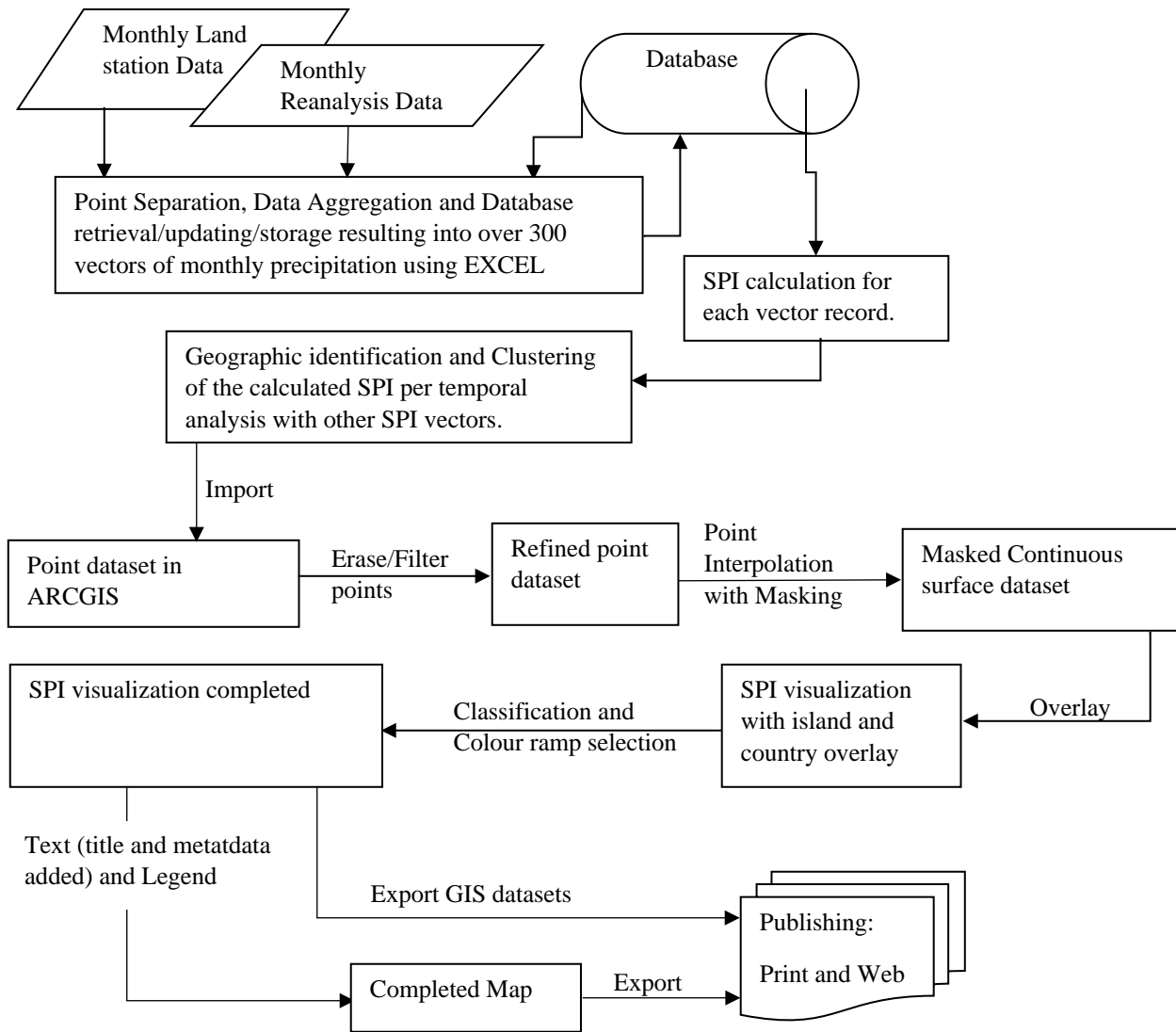


Figure 7: Operator Processed Analytical/Digital Workflow for SPI Modelling

3. An Analytical Cartographic Approach

Using the premise that the cartographic model can be applied between projects, as suggested by Brus, Dobešová, Kanok, International Conference on Intelligent and Collaborative [1], the authors sought to improve on the manual/digital model, previously described. The approach considered that there were two components to the workflow. Firstly, an analytical component that processes precipitation data to obtain SPI trends, and secondly the mapping component. To accomplish the first task, the authors used the free software environment for statistical computing and graphics R. This was preferred as R is widely used in climate modelling and packages for SPI exists. However,



although R does offer some GIS capability it is not considered on par to the mapping functionality that is capable using specialised geospatial software such as ArcGIS. This required an artistic view of map-making as is later discussed.

3.1 Analytical Computations

R was used for the analytical processing of the environmental data and extracting the SPIs. Unlike working with smaller datasets and building a database for processing, R can easily handle the datasets extracted from the source databases. The reanalysis downloaded data is in NETCDF format which can be easily converted to ASCII formats using existing R packages [7]; separated into point vectors based on the geographical locations; and processed for SPI computations using simple routines and available packages [8]. A workflow of the algorithm developed is shown in Fig. 2.

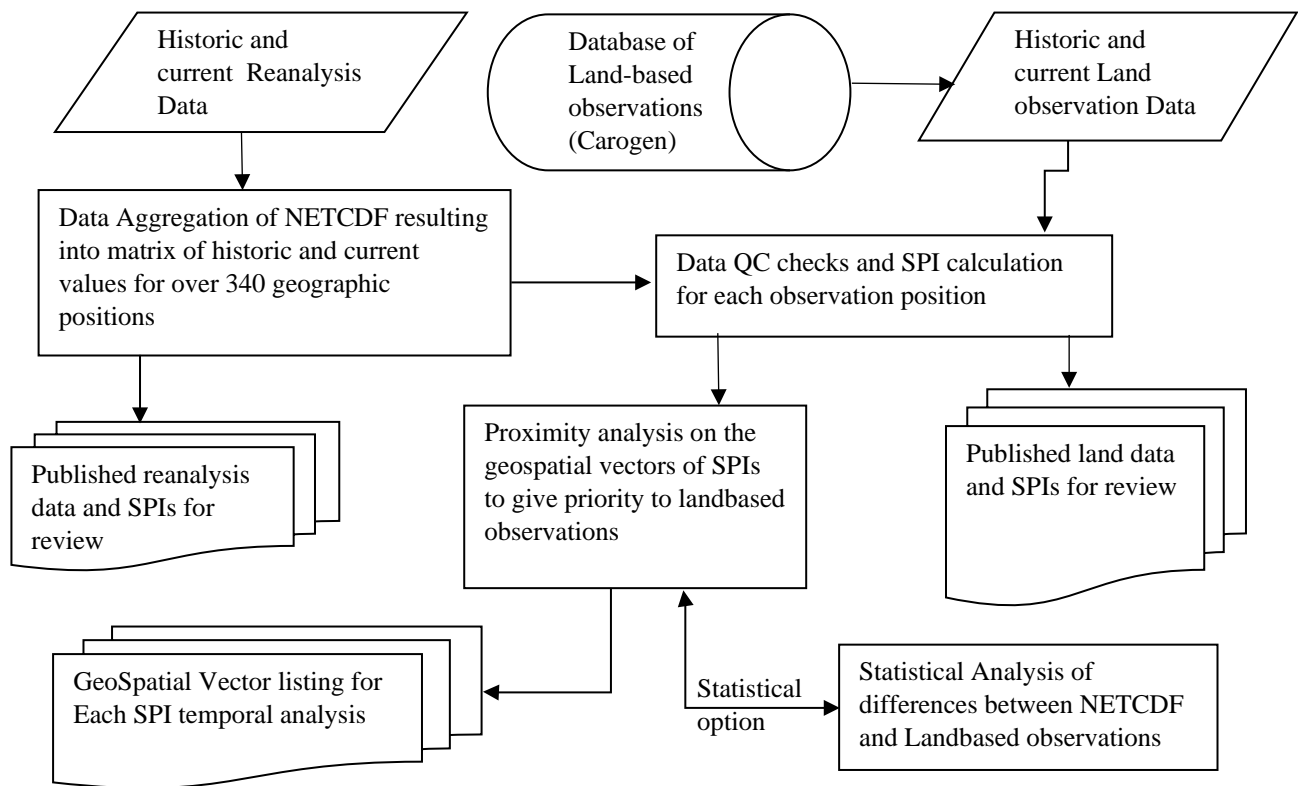


Figure 8: Automatic Analytical Processing Model for SPI Computations Using R

Unlike manual processing of the data, the algorithm did all the analysis automatically once the datasets were defined. Previously aggregating of data was done on a monthly basis and saved to a database. In contrast, the algorithm reads both the reanalysis data and land-based data as complete datasets, working with as close to source data as possible. This also made it easier to expand the



geographic extent of the analysis and apply statistical based quality checks to determine the bias of the reanalysis data, if any, as compared to the land-based observations. To ensure that transparency was not lost, data from each stage were exported in appropriate formats for readings and/or database usage. Although SPI packages are available in R, an option for using an external SPI calculator was included to ensure consistency with the previous calculations and maps. The outcome of the analytical analysis is a geospatial vector of SPI values for each of the temporal analysis period: monthly, 3 months, 6 months, 12 months, and 24 months. In addition to these, the algorithm was designed to calculate SPI differences between consecutive SPI analysis months, and precipitation averages, totals, and percentages.

3.2 Analytical Mapping

While R has graphic capabilities it is not designed for cartographic quality products such as those produced by ArcMAP. As such, the authors considered the “artistic” nature of cartography and used image manipulation based on a “Drag-and-drop pasting” methodology as suggested by Jia, Sun, Tang and Shum [9] for image composition. This was considered appropriate because there existed previous cartographic products in the form of a GIS methodology and maps. Using the existing cartographic rules (based on the ArcGIS model), the analysis applied developed methodology as outlined in Fig. 3. The SPI point listing created from the analytical components is imported into the developed mapping component. The component handles the neighbouring operations of interpolation of the geospatial point vector to create a continuous surface; categorization of SPI values based on the legend classification bin categories, generation and assignment of colours based on their Red, Green, Blue (RGB) values as used in predetermined/pre-designed legend templates; masking/clipping to define the area of interest and define spatial extents; finally, overlay a country boundary vector dataset, add title and other texts, and create map layouts.

While R does have surface/point interpolation functionality, the authors decided to use the interpolation functionality with the Generic Mapping Tools (GMT) freely available software [10, 11]. The functionality of the GMT tools allowed for additional cleaning up and aggregation of points. The algorithm calls the GMT program, and interpolates the points listing exported out of the analytical computational component as an ASCII listing. The parameters for interpolation were executed as preconfigured .bat files that included the interpolation interval.

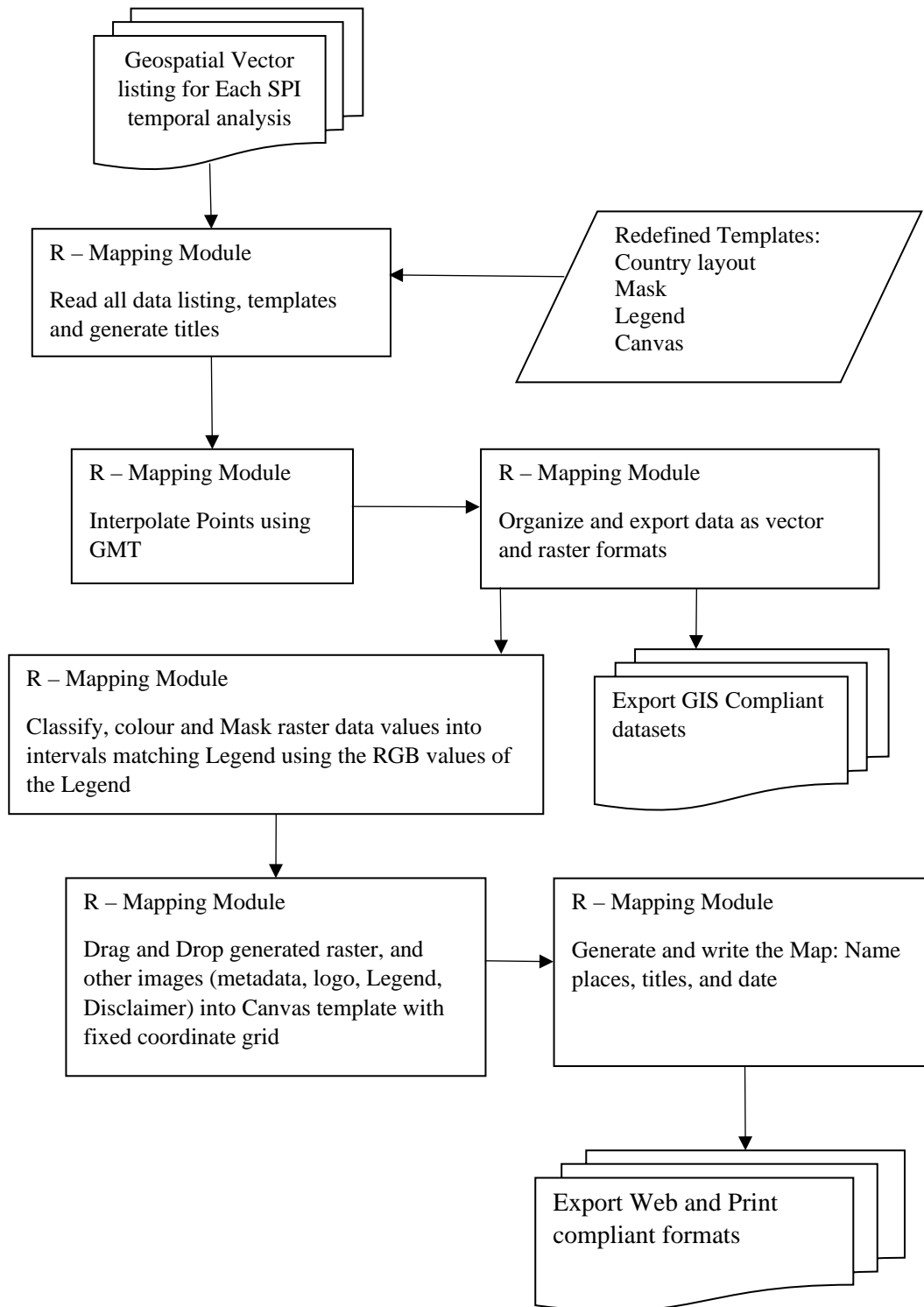


Figure 9: Automatic Analytical Processing of Processed Data for SPI Modelling using GMT and R



3.3 User Friendliness

R fundamentally uses a command line interface, that may be challenging for basic operators. In order to ensure a short learning curve and that users do not need to modify the algorithm directly, a Graphic User Interface (GUI) was implemented using widgets. In the R environment, this was implemented using a basic set of tools in the form of the gWidgets package [12]. This allowed the algorithm to be wrapped within the Widgets GUI. Using this simple approach, the operator runs the algorithm which opens the GUI and awaits confirmation from the operator as to the datasets, and dates. During the execution of the algorithm, various confirmation boxes update the operator of the processing and notifies them when completed. Default parameters, such as dates, QC threshold, can be easily changed within the GUI.

4. Results

In terms of computational requirements, the developed methodology involved much more processing. Manual data processing dealt with daily reanalysis data for the month in consideration, and updated a localized database. In contrast, the analytical solution worked with the wider temporal period of the daily reanalysis data encompassing the historic period of analysis. This meant that the input data was much larger and needed to be aggregated from daily to monthly totals; hence requiring more processing. Notwithstanding the file size, analytically, the processing time of computing SPIs was a matter of minutes (10 to 30 mins, depending on computational power), compared to the manual sorting and processing which took about 1½ day. Similarly, processing of land-based observations read a copy of the database and processed the data in a few minutes.

The operators' use of ArcGIS for mapping, typically required half-a-day (½ day) for the analysis and production of six (6) maps. The developed analytical mapping solution was significantly shorter taking less than one (1) hour to compute the six (6) SPI maps in addition to other new map products that compared the change in SPIs from the previous month. Another advantage of the developed system is that the map layouts were consistent and the processing adequately removed reanalysis points that were either spatially and/or numerically inconsistent to the land-based observation. Table 1 summarizes the differences and inherent problems associated with the methodologies. Figure 4 shows a sample SPI map made using the ArcGIS method, while Fig. 5 shows a sample from the developed methodology.



Table 3: Comparison of results.

Category	Original Method	Analytical Solution	Remarks
Data size and storage	Operator worked with subsets of the data and smaller databases	Processing done using original dataset	The original reanalysis datasets are in daily totals and processed over the full analysis period of over 18 years.
Time	1½ - 2 days	10 – 30 mins	Maps can now be produced 75% faster.
Productivity	Operator-intensive	Computer processing intensive	Operators can focus on alternative tasks.
Number of products	6 maps	13+ maps	Additional maps are created with the analytical solution. Including differences between SPIs and rainfall percentages. Each published in 3 image formats and a vector format.
Expertise required	Basic MS Excel, Windows and ArcGIS skills are required.	Basic MS Windows skills are required.	With the use of the GUI, very little skills are required with the analytical solutions.
QA/QC	Checks done by the operator. Earlier interception of qualitative related problems of the data.	Checks done by the algorithm. Better organization of data and quantitative checks.	The algorithm does better checks on the integrity of the source data and differences between satellite and land based data. However, qualitative checks are done when the maps are published.
Consistency	More care needed when working with reanalysis subsets. Higher levels of skills required to ensure consistency.	Highly effective but still dependent on quality of reanalysis data.	Consistency of the reanalysis data can be guaranteed as data is consistently updated within the first few weeks.

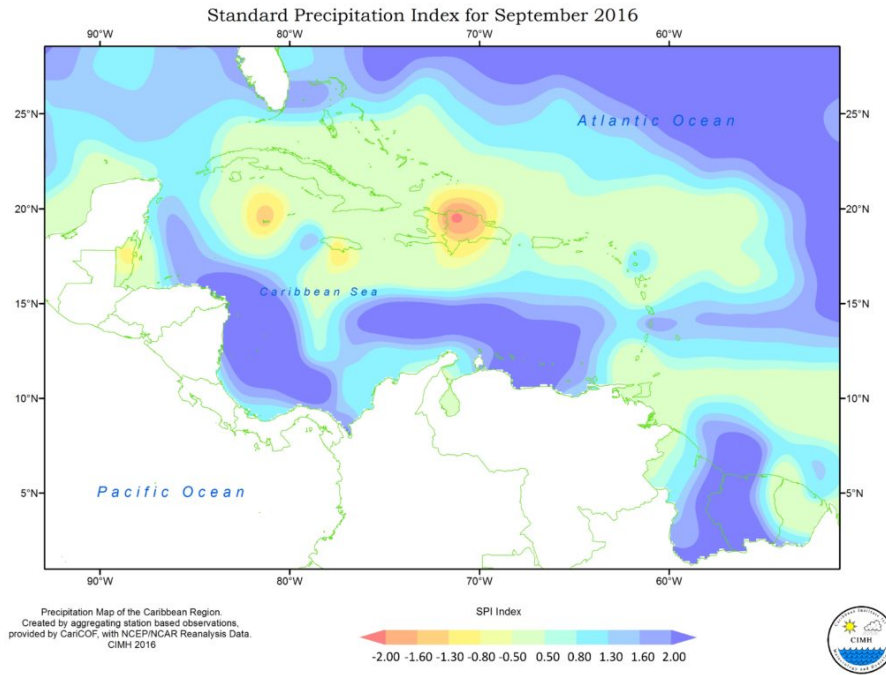


Figure 10: Map produced using the ArcMap Software
[<https://rcc.cimh.edu.bb/spi-monitor-september-2016/>]

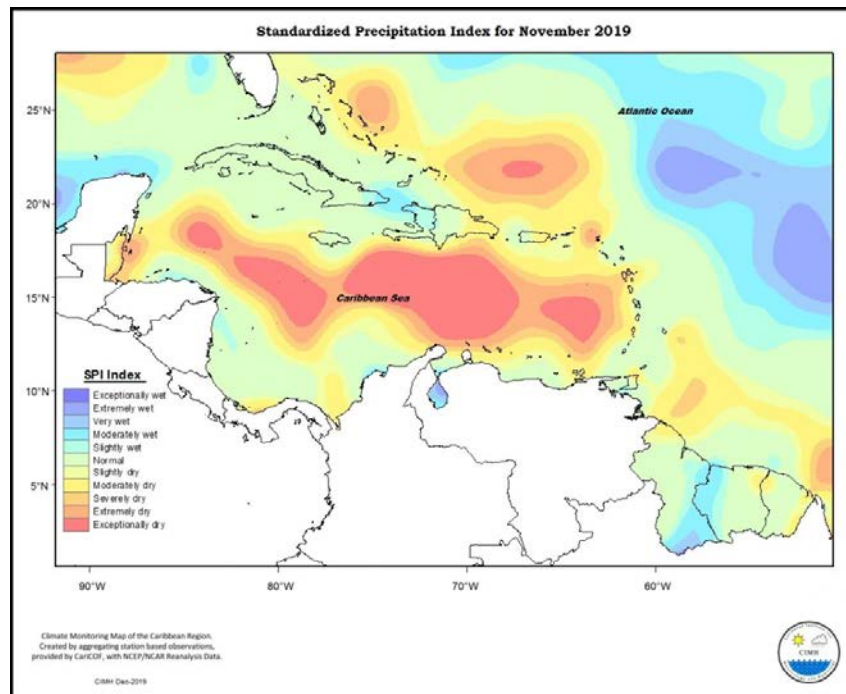


Figure 11: Map automatically produced using the developed methodology in R
[<https://rcc.cimh.edu.bb/spi-monitor-november-2019/>]



5. Discussion and Conclusion

With the use of R and a GUI, the productivity was high with a low learning curve. Overall, the computational and mapping time was reduced from 2 days to ½ hour. The maps were highly repeatable and reproducible using the automated system. In addition, there were improvements in data management, with all processes now done automatically using the full downloaded dataset, eliminating errors caused by data handling and sorting. These included, placing data in the wrong column and not taking database changes into consideration. The latter was found to be a significant source of error in the old methodology, as reanalysis data are numerical estimates for the first few weeks and improved as more and more data is processed. The old methodology of working with the monthly data did not capture such changes. In addition, the fully integrated automated system eliminated much of the subjectivity associated with making maps, as such increasing the reliability of the information depicted, with any remaining uncertainty associated with the data rather than the cartographer.

The developed system significantly expanded the products exported. Over 200 files consisting of database ready formats, shapefile point vector datasets, and maps were exported as GeoTiff raster images and publishing formats in the form of .jpg, .esp, and .png. In addition to the SPIs, the SPIs differences, Totals per Month, and Totals per Year were also computed. With a good PC/processor these can be processed in less than 30 minutes. The methodology showed that maps produced by the analytical system can be easily manipulated to be of high cartographic standards, comparable to the GIS solution using commercial software.

While the R environment was shown to be fully capable for analytical cartographic modelling, it is important to note that this was based on an existing model of the previous work done using the ArcGIS software, and as such, cartographic rules were already considered. In addition, R algorithm involved over 2000 command lines in addition to 45 packages. This has been found to be a drawback when the base R software and/or modules are upgraded as some calling formats may change and no longer work. Also, the use of highly specialized software within the Caribbean means that the number of people that can modify and correct the code is limited to a few professionals as compared to the broader GIS user base.

Acknowledgment

We thank the Caribbean Institute for Meteorology and Hydrology (CIMH) for the opportunity to carry out this work. This was carried out through the United States Agency for International Development's (USAID) Programme for Building Regional Climate Capacity in the Caribbean (BRCCC Programme), executed by the World Meteorological Organization (WMO), and implemented by CIMH.

References

- [1] J. Brus, Z. Dobešová, J. Kanok, N. International Conference on Intelligent, and S. Collaborative, "Utilization of Expert Systems in Thematic Cartography," pp. 285-289, 2009.
- [2] M. Monmonier, "How to Lie with Maps," 1991.



- [3] C. S. National Committee for Geography, *Glossary of technical terms in cartography / prepared on behalf of the Cartography Subcommittee of the British National Committee for Geography*, London: The Royal Society, 1966.
- [4] H. Moellering, "Analytical Cartography: Past, Present, and Future." pp. 599-614.
- [5] S. Steiniger, and E. Bocher, "An overview on current free and open source desktop GIS developments," *International Journal of Geographical Information Science*, vol. 23, no. 10, pp. 1345-1370, 2009/10/01, 2009.
- [6] M. M. Venigalla, and B. H. Baik, "GIS-Based Engineering Management Service Functions: Taking GIS beyond Mapping for Municipal Governments," *Journal of Computing in Civil Engineering*, vol. 21, no. 5, pp. 331-342, 2007.
- [7] P. Michna, and M. Woods, "RNetCDF - A Package for Reading and Writing NetCDF Datasets," *The R Journal*, vol. 5, no. 2, pp. 29-36, 2013, 2013.
- [8] S. Čadro, and M. Uzunović, "HOW TO USE: Package 'SPEI' For BASIC CALCULATIONS," 2013].
- [9] J. Jia, J. Sun, C.-K. Tang, and H.-Y. Shum, *Drag-and-drop pasting: Association for Computing Machinery*, 2006.
- [10] P. Wessel, and D. Bercovici, "Interpolation with Splines in Tension: A Green's Function Approach," *Mathematical Geology*, vol. 30, no. 1, pp. 77-93, January 01, 1998.
- [11] P. Wessel, W. H. F. Smith, R. Scharroo, J. Luis, and F. Wobbe, "Generic Mapping Tools: Improved Version Released," *Eos, Transactions American Geophysical Union*, vol. 94, no. 45, pp. 409-410, 2013.
- [12] J. Verzani, "An Introduction to gWidgets," *The Newsletter of the R Project*, vol. 7/3, no. December 2007, pp. 61, 2007, 2007.