RESEARCH NOTE

Visualization of spread of papaya ring spot virus using "raster" package in r statistical environment

**DULEEP KUMAR SAMUEL, M. KRISHNA REDDY, SALIL JALALI,
H.C. REDDY and M. MANASA**

Virology Laboratory, Division of Plant Pathology, Indian Institute of Horticultural Research, Bangalore - 560 089, India
Email:dksamuel@gmail.com

Papaya (*Carica papaya* L.) is an important tropical fruit crop with an increasing market as a health food. India is the leading papya producer. The Papaya Ring Spot Virus (PRSV) is emerging as a major limiting factor in the production of Papaya. The PRSV causes mosaic on the leaves, leading to malformed leaves. Water soaked lesions appear on the stem, fruit and flower drop would be seen, Malformed fruits appear with characteristic rings on the fruits. Affected plants become unproductive and would die soon after. The PRSV had been extensively described (Singh, 2003 and Tripathi *et al.*, 2008) and is vectored in a non-persistent manner, by aphids *Aphis gossypii* Glover, *Myzus persicae* (Sulzer) and *A. craccivora* Koch (Kalleshwaraswamy and Kumar, 2008). The spread of PRSV would be quite rapid in the field as the rate of spread determines the economic loss caused and it is necessary to visualize this spread. An attempt was made to visualize these parameters using R software which is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS, It is available for free download at (http://www.r-project.org). It has the advantage of providing a free alternative to costly commercial software. Geographic Information Systems (GIS) can be used to effectively visualize spatial data. Raster is a package which runs under R and it was used to visualize the spread through charts. raster uses a spatial (geographic) data structure that divides a region into rectangles called as 'cells' (or 'pixels') that can store one or more values for each of these cells. Such a data structure could also referred to as a 'grid' and is often contrasted with 'vector' data that is used to represent points, lines, and polygons. The raster package has functions for creating, reading, manipulating, and writing raster data. The package provides, general raster data manipulation functions that can easily be used to develop more specific functions. It also implements raster algebra and most functions for raster data manipulation that are common in GIS systems.

The PRSV disease incidence was recorded on an increasing scale of 1-9 as was present in an 11x43 plot of papaya field infected with PRSV, at various time points. The data were entered in a Microsoft Excel 2013 worksheet in a r x c matrix. The disease incidence was recorded at seven different time periods *viz* 3 Jan 2012, 18 Jan 2012, 25 Jan 2012, 15 Feb 2012, 23 Feb 2012, 1 March 2012, 8 March 2012. Each time period was recorded in a separate work sheet. The changeover in disease incidence was recorded by subtracting the disease incidence value at the (n^{th}) time point from the $(n\text{-}1^{th}$ time point), this difference was saved in a separate sheet and indicated the changeover of disease incidence at that time point. All the data recorded Microsoft Excel 2013 worksheet files were finally saved as *.csv (coma separated files) as R and the program raster can read *.csv files natively.

Using a Personal Computer with Intel(R) Core(TM) i7-2600K CPU @ 3.40GHz, with 32 GB RAM, a NVIDIA GeForce GTX 560 Ti video card, with 3 x 2 TB drives and running Windows 8.1 64 bits the collected data was analyzed. First "R version 3.02 64 bits" was installed and then the package raster was installed from CRAN. The charts were made by issuing the following commands.

```
x <- read.csv (file.choose()) # helps to choose the data file
require(raster) # raster is the package for the analyses
b <- as.matrix(x) # changes the data to a matrix
r <- raster( b ) # raster computes the figure matrix
plot( r, axes=F ) # plots and the figure to the screen
```

Command x <- read.csv would assign the csv file values to the variable x, and the command file.choose (), would help to choose the correct csv file using the windows explorer as the file manager.

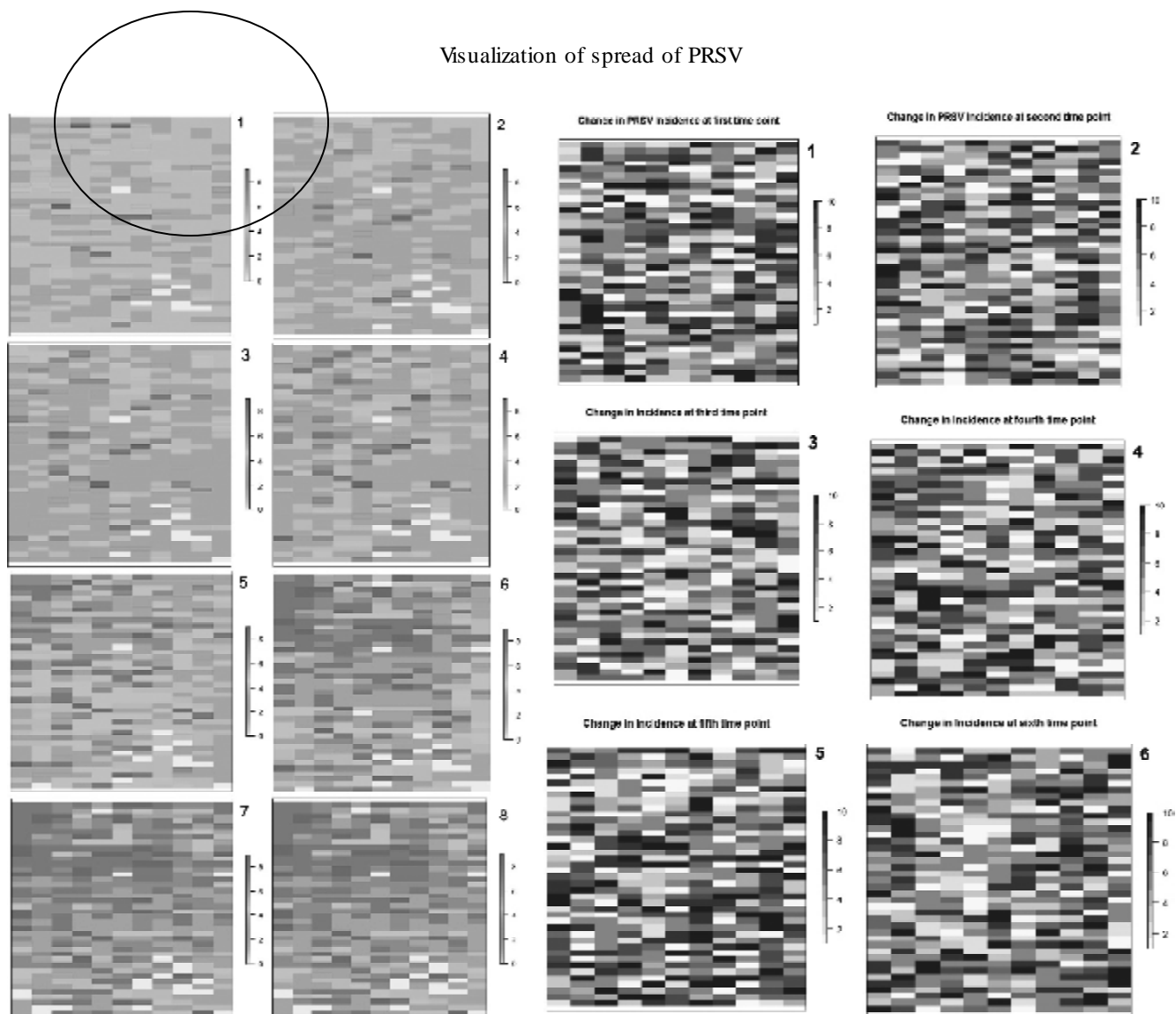Require (raster) command would load raster

Fig. 1. PRSV infection at 7 time points, (sub charts 7 and 8 are identical and are the final measured values) Right; Rate of change of PRSV infection at 6 time points. Time points are 3 Jan 2012, 18 Jan 2012, 25 Jan 2012, 15 Feb 2012, 23 Feb 2012, 1 March 2012, 8 March 2012 have been indicated by numerals (1-7) at top left position in each sub chart.

package, and the following commands b <- as.matrix(x), would assign the variable (b) as a matrix of (x), and the variable (r) as a raster object of (b) (which itself is a matrix of (x)), Plot (r, axes = F) command would plot the raster object (r), without showing the axes. The above mentioned procedure was repeated for each spread sheet containing the disease incidence and the changeover in disease incidence spread sheets.

The results in the form of color coded graphs revealed that, PRSV spread in the field was non uniform and there was an increase in the rate of spread in certain pockets seen across the field (Fig.1). The PRSV spread was observed to originate from the lower right corner which might have been due to prevailing wind direction. No gradient type of spread as well as the infected plant

to non-infected plant spread patterns was also not seen. This in totality meant that PRSV infections were completely random. This randomness could be due to random alignment of aphid vectors and/or PRSV infection was carried over from nursery. In view of the PRSV spread as visualized, to manage PRSV it would be very necessary that initial infections would have to be contained as they were seen to be more important as compared to lateral infections. Hence use of papaya seedlings grown in virus free conditions would be beneficial. The rate of PRSV infection increased after the 3rd time point and stabilized after the 6th time point at which nearly 85 % of the plant got infected. Lecoustre *et al.* (1989) analyzed and mapped the spatial spread of African cassava mosaic virus using geo statistics and the kriging technique and recorded that geostatistics can be

applied to other viral and non-viral plant diseases. Moreno *et al* (2007) analyzed the temporal and spatial patterns of spread of Lettuce Mosaic Virus which is transmitted like PRSV in a semi persistent manner by aphid vectors. Cumulative spatial data for infected plants at different growth stages were analysed using spatial analysis by distance indices. The results that secondary cycle of spread occurred when no colonising aphid species landed on the primary infected plants (probably coming from infected seed) and move to adjacent plants before leaving the crop. The role of weeds growing close to lettuce fields as potential inoculum sources of virus and the aphid species most likely involved in the transmission of LMV were also identified. The nature and the rate of spread determine the economic loss and it is necessary to visualize them. This charting protocol using the free R software environment will help to define the effects of interventions on the spread of a disease in a plant ecosystem. Apart from viral diseases, any character (yield, height, fertilizer response, pest and disease incidence can be similarly mapped using the simple "R: code"

ACKNOWLEDGEMENTS

REFERENCES

Kalleshwaraswamy, C. M., Kumar, N. K. 2008, Transmission efficiency of Papaya ring spot virus by three aphid species. *Phytopathology*, **98**(5): 541-546.

Lecoustre, R., Fargette, D., Fauquet, C., and de Reffye, P. 1989. Analysis and mapping of the spatial spread of African cassava mosaic virus using geostatistics and the kriging technique. *Phytopathology*, **79**: 913-920.

Moreno, A. M., Nebreda, B. M., Diaz, M. Garcý´a, F. Salas and Fereres, A. 2007. Temporal and spatial spread of Lettuce mosaic virus in lettuce crops in central Spain: factors involved in Lettuce mosaic virus epidemics, *Annals of Applied Biology,* **150**. 351-360.

Singh, S. J. 2003. Virus and phytoplasma disease of papaya, passion fruit and pineapple. Kalyani Publishers Ludhiana, pp. 147.

Tripathi, S., Suzuki, J. N. Y., Ferreira, S. A., and Gonsalves, D. 2008. Papaya ringspot virus-P: Characteristics, pathogenicity, sequence variability and control. *Molecular Plant Pathology,* **9**(3): 269-280.