## GISVET 2004 – The University of Guelph

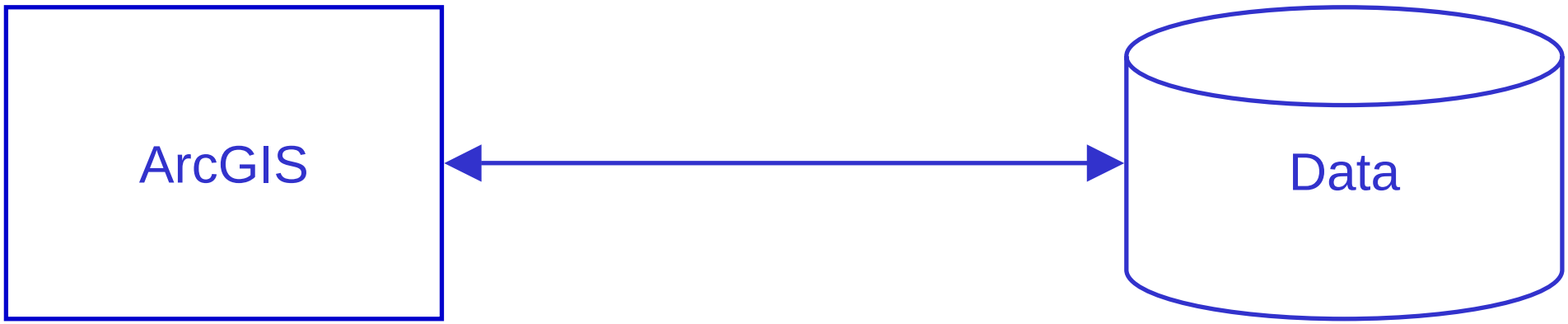
# Linking R and ArcGIS - Developing a Spatial Statistical Toolkit for Epidemiologists

N. G. Tait[1], P. A. Durr[1] and P. Zheng[2]

[1] Veterinary Laboratories Agency, UK
[2] Lancaster University, UK

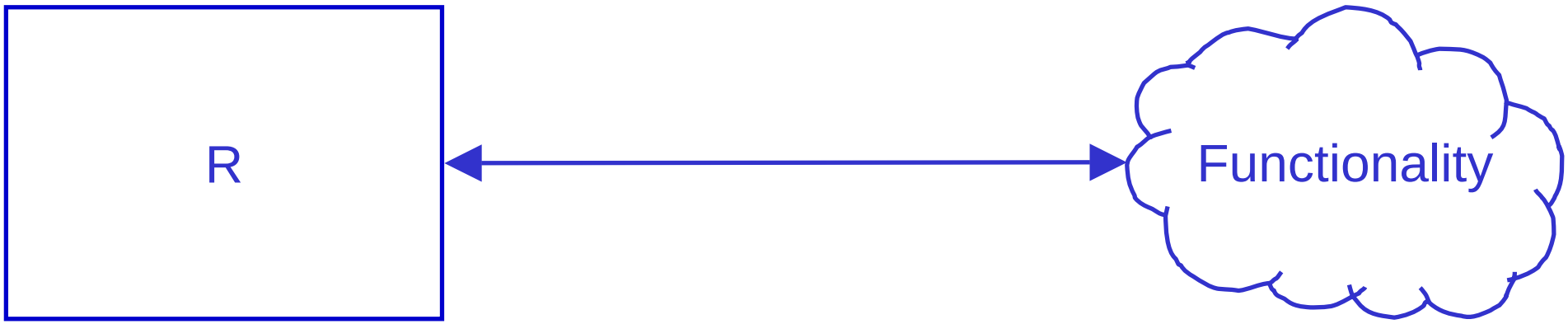
Nigel G Tait, GIS Developer, VLA - Weybridge, New Haw, Addlestone, Surrey, UK. t: +44 (0)1932 357784 e: n.tait@vla.defra.gsi.gov.uk

- **Problem Definition**



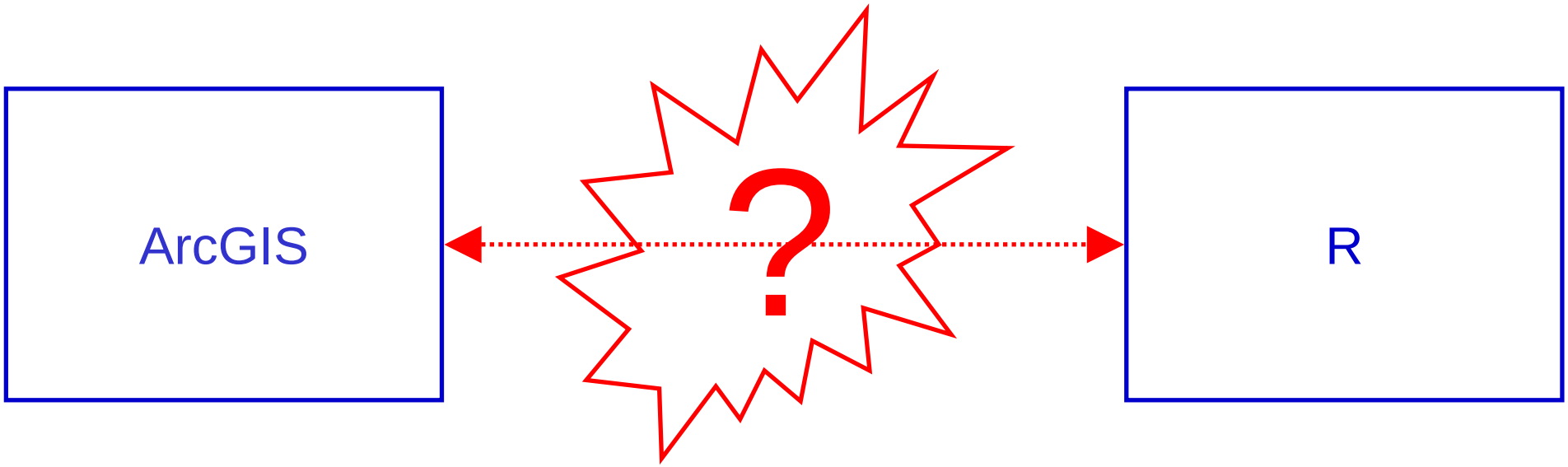

- ArcGIS - Commercial desktop GIS package
- Windows GUI
- Object model
- Poor spatial statistical functionality in extensions
- Used to store and visualise disease data

## Problem Definition (Cont…)




- R - Freeware statistical and graphics package
- Command line GUI
- No object model
- Rich spatial statistical functionality in packages
- Used to analyse disease data

- **Aims and Objectives**




- To seamlessly expose the functionality developed in the R statistical package to the data stored in the ArcGIS environment
- Enable spatial statistical analysis of disease data within a user friendly desktop GIS system

- **Coupling Methodology**


- Loose (manual) coupling
- Access to functionality via manual transfer of data files
- No development overhead
- Time consuming and error prone

# Coupling Methodology (Cont…)


- Loose (automated) coupling
- Access to functionality via automatic transfer of data files via pre- and post- processor routines
- Little development overhead
- Time consuming

# Coupling Methodology (Cont…)


- Tight (full) coupling
- Embedding functionality within the client GIS via macro and scripting languages
- Languages not powerful enough to implement complex algorithms

# Coupling Methodology (Cont…)


- Tight (close) coupling
- Development of functionality within compiled DLLs
- Enabled use of more advanced programming languages
- Seamless integration of functionality

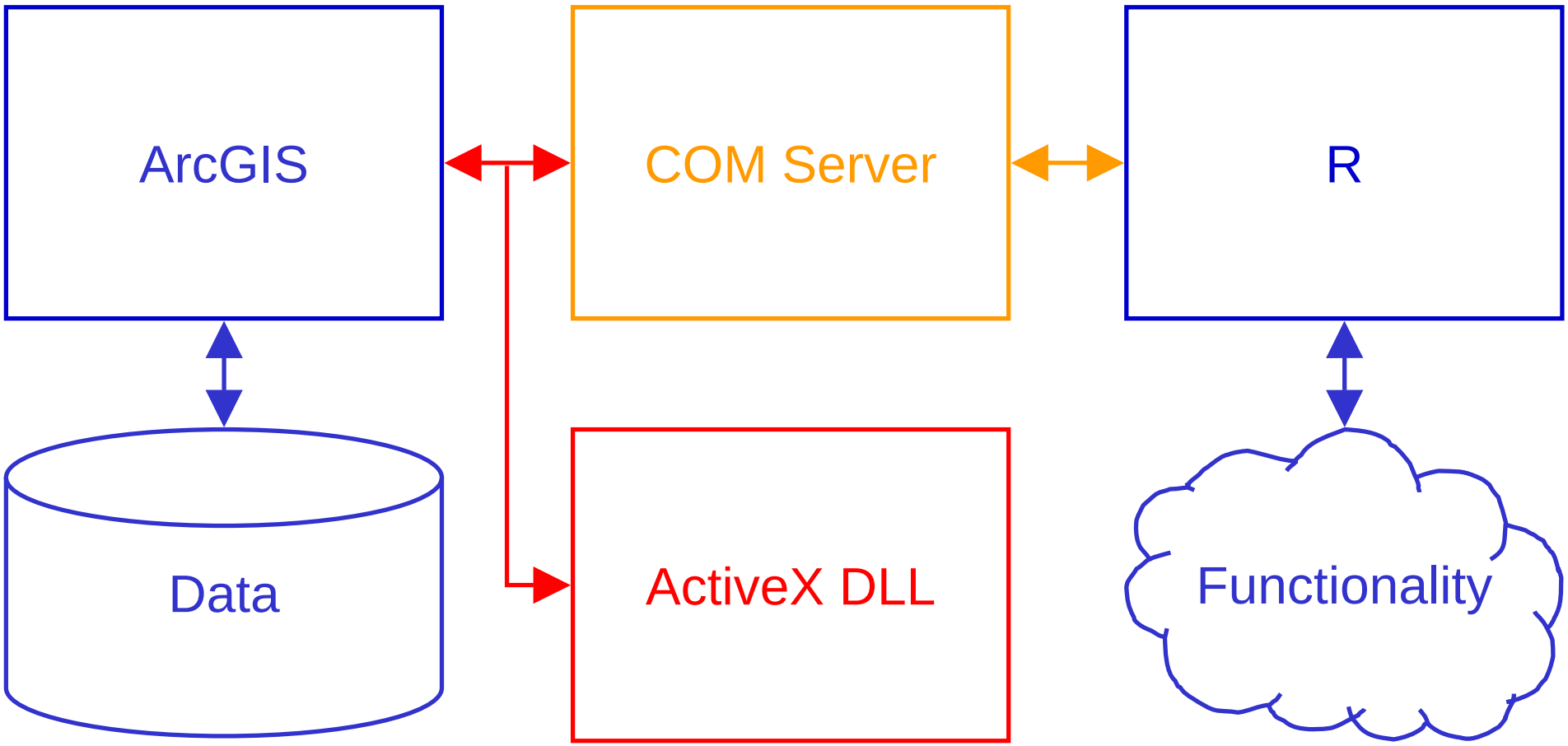
## Coupling Methodology (Cont…)


- Tight (shared) coupling
- Access to functionality through links between Microsoft Windows applications
- Access to SQL databases, non SQL data and objects
- Seamless integration of functionality

- **Conceptual Model**


- COM server provides functions to send / retrieve data and evaluate statements in R interpreter space
- Developed for Microsoft Excel
- ArcGIS can access R functionality via COM server application

# Conceptual Model (Cont…)



- **ActiveX DLL Component**


- VB6 ActiveX DLL - ArcGIS extension
- References to ESRI ArcObjects and COM Server libraries
- Effective link between ArcGIS and COM Server object models

# ActiveX DLL Component (Cont…)


- Provides GUI for data and parameter selection
- Validates data and parameters
- e.g. Ensure polygons have single geometry (no islands, no holes)
- e.g. Ensure points are within polygon boundary

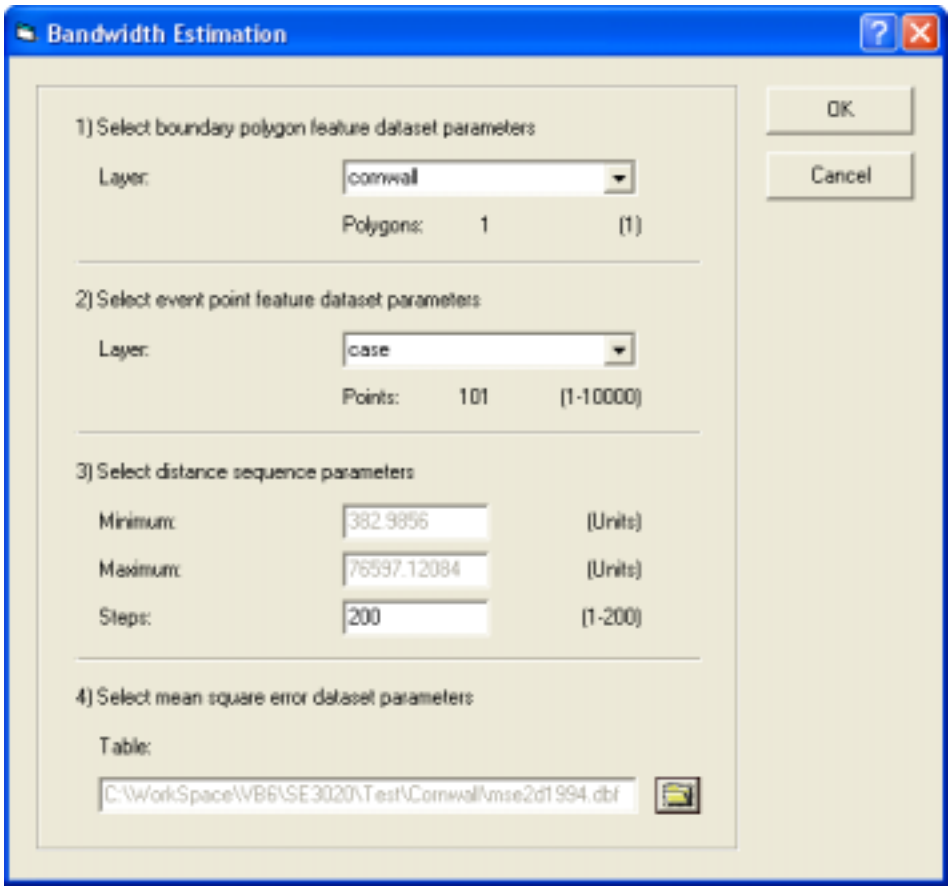
# ActiveX DLL Component (Cont…)

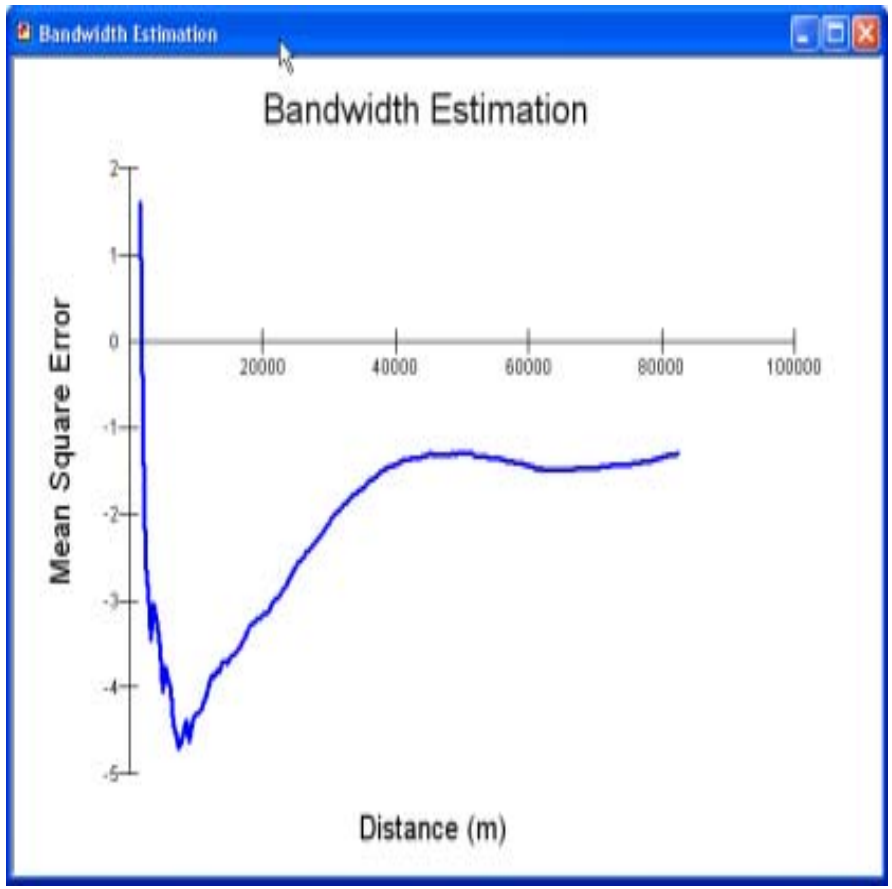

- Transforms data and parameters into structures compatible with COM Server and visa-versa
- e.g. Converts a polygon geometry to a 1 dimensional array of anticlockwise x … y vertex coordinates
- Handles errors

- **Implemented Interfaces**

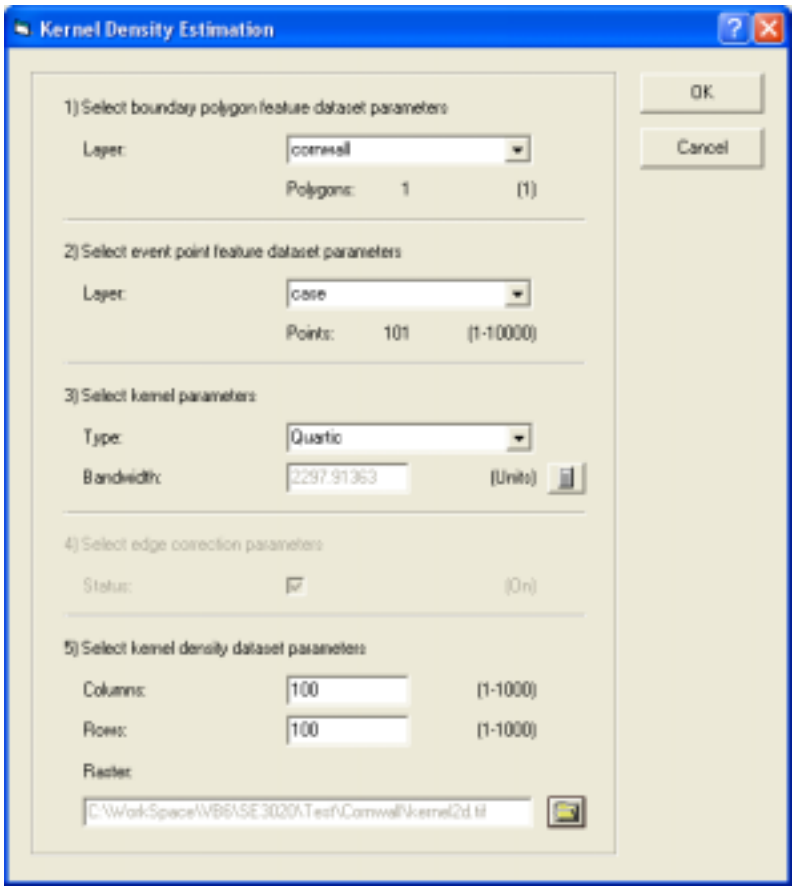
- Univariate point patterns (splancs package)
- Mean square errors (Mse2d)
- Kernel density (Kernel2d)
- K-function and K-function envelopes (Khat, KenvCsr, KenvLabel and KenvTor )
- L-function and L-function envelopes (Lhat, LenvCsr, LenvTor)
- Multivariate point patterns (tb package)
- Cross-validated log-likelihood function (Lc)
- Kernel density (Lambdahat)
- Type specific probability (Phat)

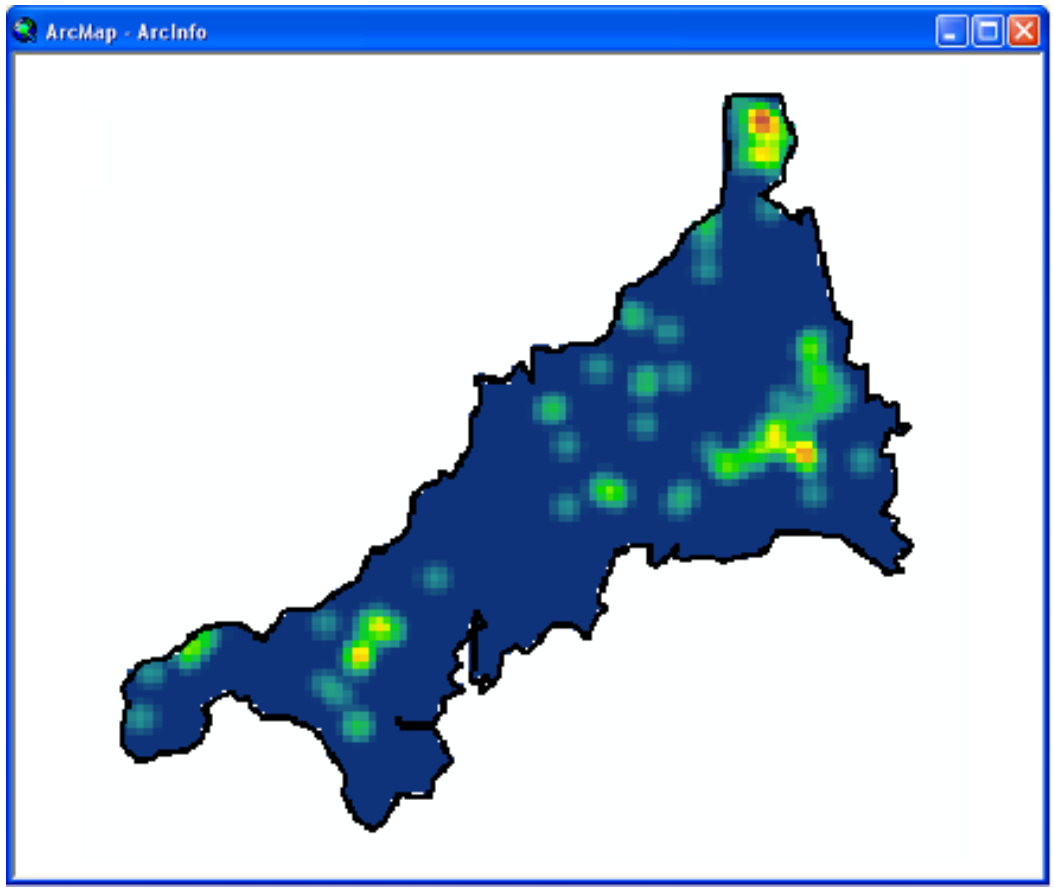
# Implemented Interfaces (Cont…)



# Implemented Interfaces (Cont…)



## Conclusions

- Proved that its is possible to link ArcGIS and R, this being one of the most successful attempts documented
- However there are issues to resolve
- VB Debugger runs ActiveX DLL out of process from ArcMap, placing a Debug DLL in process space instead. Therefore marshalling fails with unsupported interfaces
- No method of debugging when program control is passed to R, therefore data needs to be strictly validated before being passed

## Conclusions (Cont…)

- Data structures differ between R packages
- e.g. splancs package polygon can be either open or closed while tb package polygon is open with anticlockwise orientation of vertices
- No method of process progress reporting when program control is passed to R, therefore program may appear to have crashed when processing
- e.g. K function envelopes may take long time depending on the complexity of the geometry
- Final build released as freeware as part of the GISVET initiative

**Thank you for your attention…**