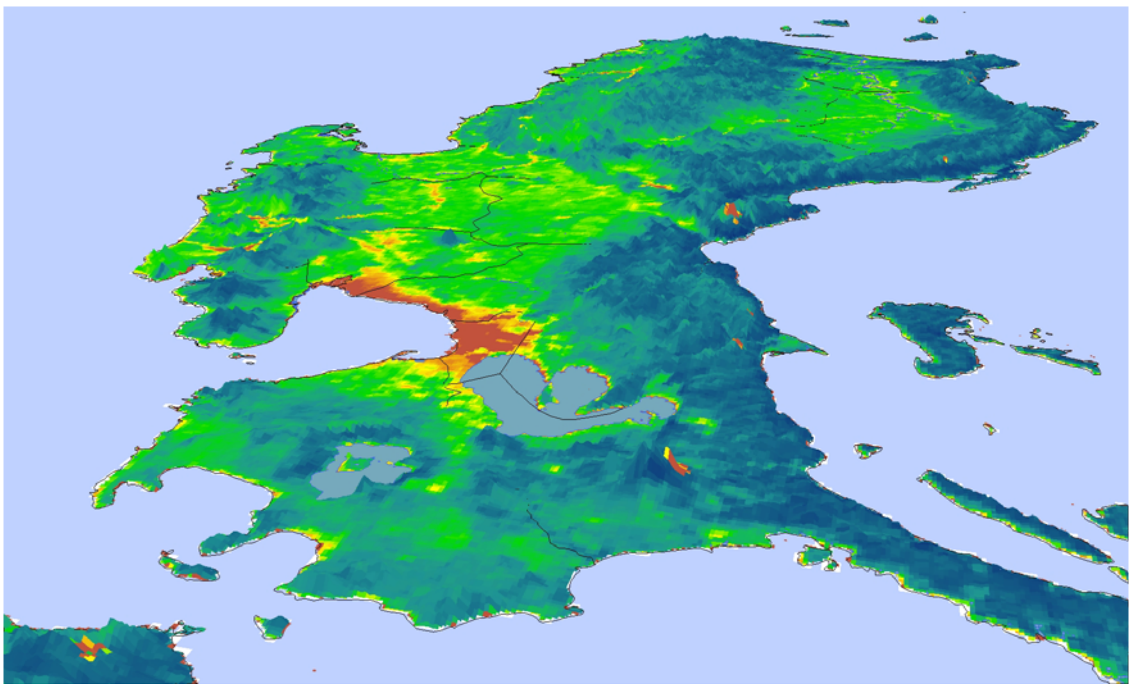
Food and Agriculture Organisation of the United Nations
Viale delle Termi di Caracalla, 00100 Rome, Italy

CONSULTANT'S REPORT

SPATIAL ANALYSIS OF ANIMAL DISEASE DISTRIBUTION IN THE PHILIPPINES

STEP TWO: IMAGE ACQUISITION AND PROCESSING AND RECOMMENDATIONS FOR FUTHER INVESTIGATION



prepared for the

Environmental Animal Health Management Initiative

by

William Wint
Environmental Research Group Oxford
P.O. Box 346,
OXFORD, OX1 3QE, UK
Tel: +44 1865 271257 / +44 1584 819574
Fax: +44 1865 310447 / +44 1584 819574

Andy Tatem
Spatial Ecology and Epidemiology Group
Oxford University Department of Zoology
South Parks Road
OXFORD, OX1 3PS
Tel: +44 1865 271262

May 2007

The cover illustration is a MODIS 1km resolution satellite image depicting the mean Enhanced Vegetation Index (EVI) product for most of Luzon Island, northern Philippines, draped over a 3D Digital Elevation Map from the Shuttle Radar Topography Mission

SUMMARY AND CONCLUSIONS

The long term development goal of the Environmental Animal Health Management Initiative (EAHMI) is to promote sustainable agriculture and rural development, through environmental animal health management for enhanced smallholder livestock production in the Philippines.

Establishing a Geographical Information System (GIS) is an essential element in creating a credible suite of spatial analyses envisaged by EAHMI, which depends on access to livestock and disease information that is substantially more detailed than the provincial level data currently available. This involves the creation of detailed maps of animals and their diseases.

This report documents the second in a series of innovative measures identified as necessary to implement the analyses required, and follows on from recommendations of the initial consultancy (Wint & Tatem, 2007). First amongst these was the production of remotely sensed imagery to provide the core of a data archive needed to enhance existing livestock and disease maps using established statistical distribution modelling techniques.

The sequential steps involved in the production of a set of these environmental datasets derived from National Aeronautics and Space Administration's (NASA) MODerate-resolution Imaging Spectroradiometer (MODIS) imagery are described in some detail for the record and future reference.

These data are a global first, and have been delivered and installed within BAI. The information collected by EAHMI and collaborators has been assessed in relation to the likely needs of livestock and disease mapping. The ensuing steps required to implement statistical distribution modelling for selected livestock species (carabao, cattle, pigs, and horses) and diseases (Foot and Mouth Disease, Surra, Fasciolosis, and Hemorrhagic septicaemia) are described in some depth.

In the light of the skills available in the Philippines, it is strongly recommended that as much of the data preparation and processing is undertaken by EAHMI and its collaborators, and that the involvement of costly external analysts be restricted to interactive disease modelling procedures, reporting and delivery. A series of detailed recommendations and data requirements are provided, together with a provisional schedule.

Acknowledgements

The imagery provided with this document was processed and extracted in close collaboration with Professor David Rogers, Drs Simon Hay and Jorn Sharlemann and David Benz of the Spatial Ecology and Epidemiology Group, Department of Zoology, Oxford University, UK, to whom the authors are most grateful. Dr David Bourn and Dr Jose Molina of FAO in the Philippines have also provided extensive advice and assistance with the content of the document, which is greatly appreciated. WW would also like to gratefully acknowledge the generous hospitality provided by all concerned at FAO EAHMI, BAR, BAI and UPLB.

Table of Contents

List of Figures and Maps

List of Tables

Introduction and Overview

The long term development goal of the Environmental Animal Health Management Initiative (EAHMI) is to promote sustainable agriculture and rural development, through environmental animal health management for enhanced smallholder livestock production in the Philippines. Three sets of objectives and corresponding activities have been identified to achieve this goal: strengthen institutional capacity; formulate strategies for enhanced smallholder production, and integrate the principles of environmental animal health management into national policy and planning objectives.

Strengthening capacity for environmental animal health management involves a variety of inter-related activities, including: assessing training needs; arranging appropriate training courses; compiling and reviewing available information; identifying gaps and alternative sources; networking and commissioning complimentary studies; and developing a geographical information system (GIS) for the livestock sector.

Establishing a livestock GIS is an essential element in creating a credible suite of spatial analyses envisaged by EAHMI, which depends on access to livestock and disease information that is substantially more detailed than the provincial level data currently available. This involves the creation of detailed maps of animals and their diseases.

This report documents the second in a series of innovative measures identified as necessary to implement the analyses required, and follows on from recommendations of the initial consultancy (Wint & Tatem, 2007). The sequential steps involved in the production of a set of environmental datasets derived from MODerate-resolution Imaging Spectroradiometer (MODIS) imagery from the international Terra satellite managed by the United States National Aeronautics and Space Administration (NASA) are described in some detail for the record and future reference.

While the focus of this exercise is the Philippines, the extraction and production of the data required was conducted in collaboration with an ongoing project that is acquiring and compiling a global dataset of MODIS imagery. This has helped to identify numerous technical issues and processing difficulties that may not have been obvious had the focus been on the production of Philippines imagery alone. This is particularly true for issues that marginally affect the Philippines coverage and might have been overlooked, but were more significant for other areas, and for which solutions had therefore to be found. As a result this Philippines database is substantially more robust than would otherwise have been the case.

The image processing chain required to provide data layers suitable for distribution modeling and analyses illustrated in Figure 1 and described in the following sections:

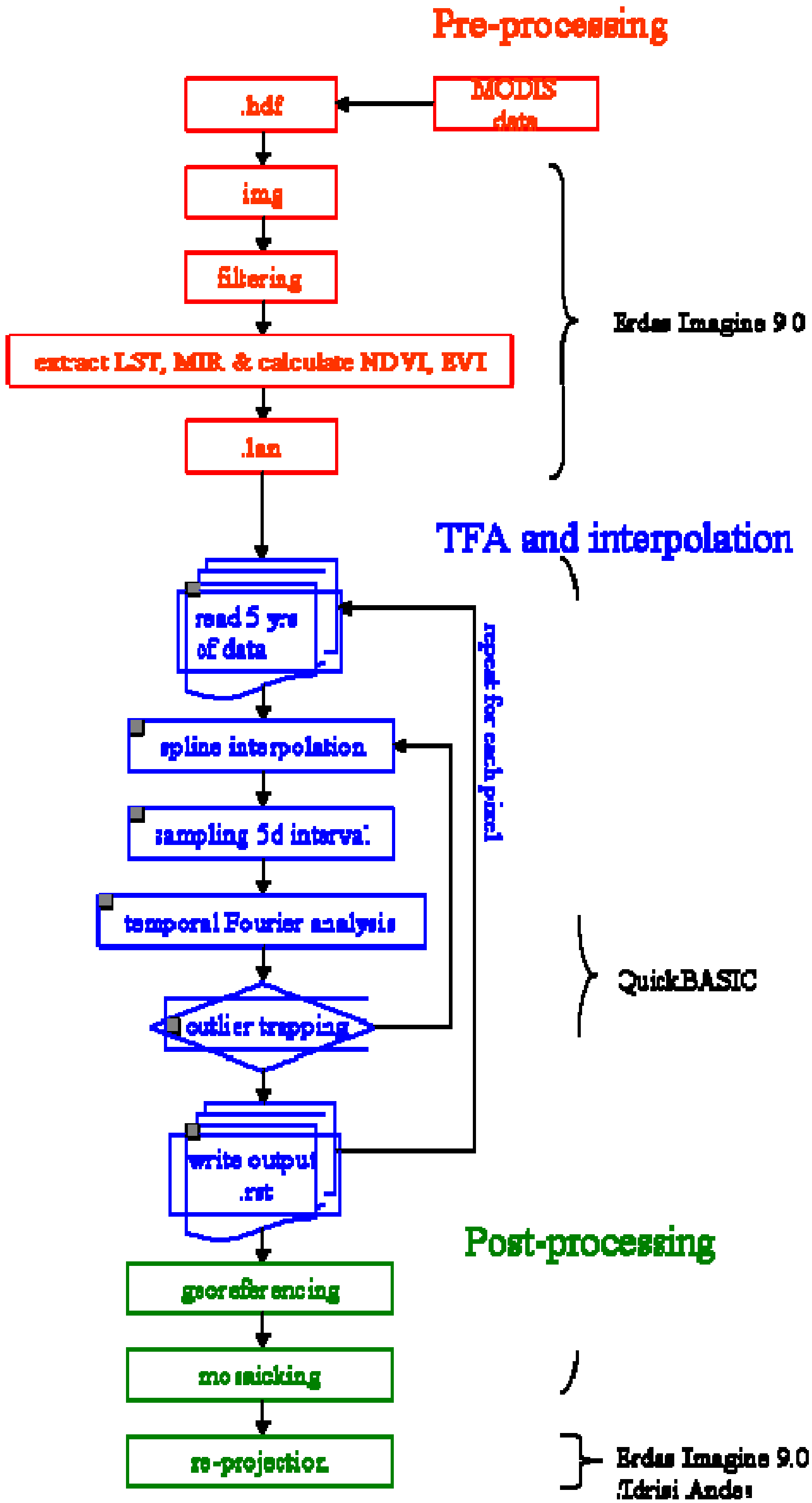
Section 2: Acquisition and download of the raw data from the supplier;

Section 3: Pre-processing, variable extraction and quality control; and

Section 4: Interpolation and Temporal Fourier Analysis (TFA) to provide potential predictor variables for distribution modeling.

Three suites of software are required to implement the complete processing sequence: ERDAS Imagine; a collection of custom written routines executed in Quick Basic: and IDRISI Andes or ESRI ArcGIS Desktop for display.

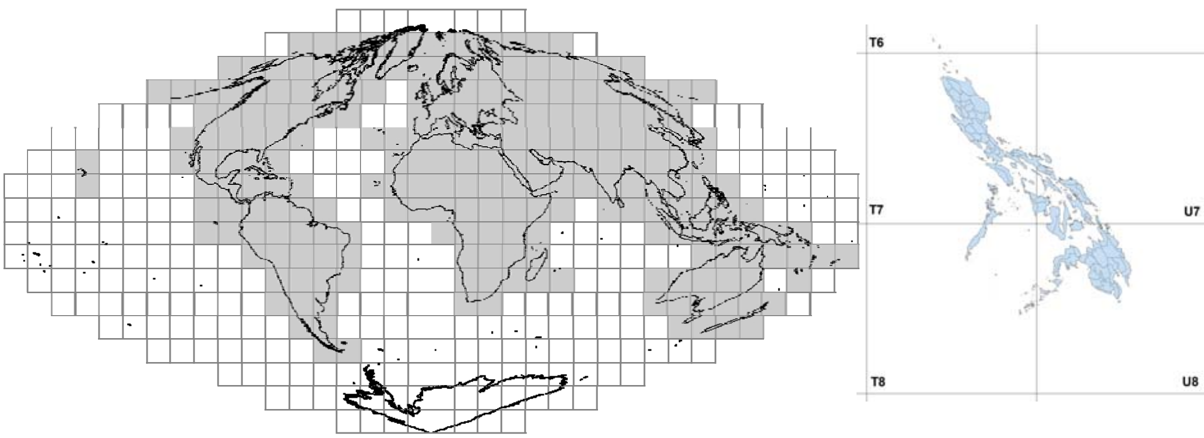
Figure 1: The MODIS Image Processing Chain



Imagery Acquisition

Time-series MODIS data at a nominal 1km x 1km (926.62543m x 926.62543m) spatial resolution from the NASA Terra satellite are available at the Earth Observing System (EOS) data gateway[1] for the period February 2000 to the present. The global MODIS dataset is provided in the MODIS Land (MODLAND) tile system in sinusoidal projection for 289 tiles, with each consisting of 1,200 x 1,200 pixels and covering 1,236,434km^2 (Figure 2).

Figure 2: The MODIS Global Dataset Tile Layout, with those Acquired for the Philippines Enlarged to the Right



Data for the year 2000 are only available from March onwards and are not readily usable for time series analyses that rely on complete annual data series. As complete data were not available for 2006, imagery for the years 2001 to 2005 was downloaded for the MODIS sinusoidal grid tiles covering all significant land areas of the Philippines (tiles t6,t7,t8,u7,u8, Figure 2), for five complete years from 1/01/2001 to 31/12/2005 for two specific datasets:

a) Day- and Night-time Land Surface Temperature[2] (dLST and nLST); and

b) Bidirectional Reflectance Distribution Function (BRDF) adjusted reflectance[3]. This dataset provides nadir reflectances for MODIS spectral bands 1-7 computed with mean solar zenith angle of the 16-day period (Schaaf, 2002). This provides reflectance values for every pixel as viewed from nadir, removing directional effects of view angle from which more reliable climatic indices can be calculated (see next section).

The full Philippines dataset amounts to somewhat more than 4,000 files, or a little over 11GB (Table 1). The complete dLST and nLST dataset for each tile consists of 46 granules (individual images for each tile) at 8-day intervals for each of 5 years, *i.e.* 230 granules for each tile, or 1,150 in total for each parameter. BRDF data for each tile comprise 23 granules at 16-day intervals for each of the 5 years, i.e. 115 granules in total for each tile, or 575 for each parameter for the Philippines coverage as a whole.

[1] http://edcimswww.cr.usgs.gov/pub/imswelcome/
[2] MODIS/Terra Land Surface Temperature/Emissivity 8-day L3 Global 1km SIN grid (MOD11A2, version 4, (Wan, 2002)) (http://lpdaac.usgs.gov/modis/mod11a2.asp)
[3] MODIS/Terra Nadir BRDF-Adjusted Reflectance 16-day L3 Global 1km SIN grid (MOD43B4, version 4, (Schaaf, 2002)) (http://lpdaac.usgs.gov/modis/mod43b4v4.asp)

MODIS file characteristics are summarised in Table 1 showing: the total number of granules (individual image files within a tile extent) available from 2001-2005 inclusive for 225 MODIS tiles; permitted minimum and maximum geophysical values; maximum permitted departure of the interpolated value from the fitted Fourier value (Section 3) for: daytime land surface temperature (dLST), nighttime LST (nLST), middle infrared reflectance (MIR), Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI).

The data for both LST and BRDF granules were compiled at source by averaging the daily imagery after initial filtering for cloud contamination – and did not use the traditional method of Maximum Value Compositing – *i.e.* selecting the maximum values to represent the selected period. This averaging will have produced smoother values with fewer high outliers.

Table 1: MODIS Characteristics

Product	No. of granules	minimum geophysical value	maximum geophysical value	maximum departure
dLST	1150	220°K	390°K	5°K
nLST	1150	220°K	390°K	5°K
MIR	575	0.0001	1	0.1
NDVI	575	-0.2	1	0.2
EVI	575	-0.2	1	0.2

1.1 Additional data

In addition to the time series of MODIS imagery, the MODLAND Digital Elevation Model (DEM) and Land/Water Mask version 4 were downloaded[4] for all tiles. Version 4, was used instead of the more recent version 5 based on MOD43B4 reflectance data (Salomon, 2006), to maintain compatibility with the LST datasets that were clipped by the version 4 land/water mask prior to production. Information on inland water and ephemeral water bodies were also extracted from the MODLAND version 4 Land/Water mask.

An up-to-date DEM from the Shuttle Radar Topography Mission (SRTM30), downgraded to 1 km resolution to match the MODIS datasets and supplemented with GTOPO30 data beyond 60°N and 58°S latitude, is included in the MODLAND Digital Elevation Model and Land/Water Mask version 5[5] and elevation was extracted for all land tiles.

Cleugh *et al.* (2007) developed an algorithm to calculate Evapotranspiration (ET) based on vapour pressure met station data and satellite-derived vegetation information to measure surface resistance. Mu *et al.* (2007) improve on this algorithm by: 1) adding spatially interpolated meteorological station derived vapour pressure deficit (VPD) and minimum temperature constraints; 2) using MODIS-derived Leaf Area Index (LAI) as a scalar to estimate canopy conductance; 3) replacing NDVI with EVI in the calculation of vegetation cover fraction: and 4) adding ground surface Evapotranspiration (ET), which is calculated as a fraction of potential evapotranspiration (PET).

The evapotranspiration (ET) product was obtained as a processed product from the Numerical Terradynamic Simulation Group at the University of Montana[6]. The imagery represents global coverage at 5km spatial resolution, with 8-day repeat for 2001-2005 inclusive.

[4] ftp://landsc1.nascom.nasa.gov/pub/outgoing/dem_sin_old
[5] ftp://landsc1.nascom.nasa.gov/pub/outgoing/c5_dem/sin
[6] http://www.ntsg.umt.edu/

Pre-processing and Quality Control

For the LST and BRDF data sets, each HDF image contained multiple “bands” of data (i.e. several images in one file) and were, therefore, imported into ERDAS Imagine 9.0 (ERDAS, 2006) to extract individual image files for each environmental variable.

Pre-processing of the dLST and nLST imagery involved the exclusion of LST values outside of the permitted range (Table 1). Pre-processing of the BRDF-adjusted imagery involved the exclusion of pixels with unreliable BRDF corrections, identified by quality control flags included within the downloaded granules. These flags were not only stored as different sections of a 64 digit number, which required custom written utilities to extract, but were ranked hierarchically and substantial effort was devoted to assessing which quality threshold should be used to discard unreliable pixel values.

From the cleaned BRDF datasets, three environmental parameters were extracted:

a) A Middle Infrared Reflectance image (MIR, MODIS band 7, 2105-2155 nm). MIR is correlated with the surface temperature, structure and water content of vegetation canopies (Boyd, 2004) and has been shown to be of use in past epidemiological studies.

b) The Normalized Difference Vegetation Index (NDVI) (Heute, 2002): $\frac{(NIR - RED)}{(NIR + RED)}$

where NIR (Near InfraRed) is MODIS band 2 and RED band 1, 841-876 nm and 620-670 nm respectively). NDVI is a measure of vegetation abundance and health and has a long history of use in remote sensing and epidemiology in particular. It does however suffer from saturation when used to monitor highly vegetated surfaces, e.g. tropical forests, and can also be adversely affected by soil background effects. As a result, it is very sensitive to changes in the drier areas, but less effective at assessing conditions in moist areas. An alternative Enhanced Vegetation Index (EVI) was therefore calculated.

c) Enhanced Vegetation Index (EVI) (Heute, 2002): $\frac{2.5 * (NIR - RED)}{((NIR + 6.0) * RED) - (7.5 * (BLUE + 1.0))}$

where BLUE is MODIS band 3, (459-479 nm). The EVI is designed to minimise the saturation and soil background effects to which NDVI is prone.

Once calculated, values that exceeded the permitted range (Table 1) of MIR, NDVI and EVI were removed. Maps of the mean values for LST, MIR, NDVI, and EVI from the entire time series are shown in Figure 3.

Although NDVI and EVI are available at higher resolution (500m and 250m), MIR and LST data are currently only available at 1km resolution. For consistency across products and to allow processing of global imagery we therefore used 1km spatial resolution data throughout. All environmental products created were converted to ERDAS LAN format, ready for further processing. These are included within the data archive provided with this report.

Figure 3: Time Series Means for LST, MIR, NDVI and EVI

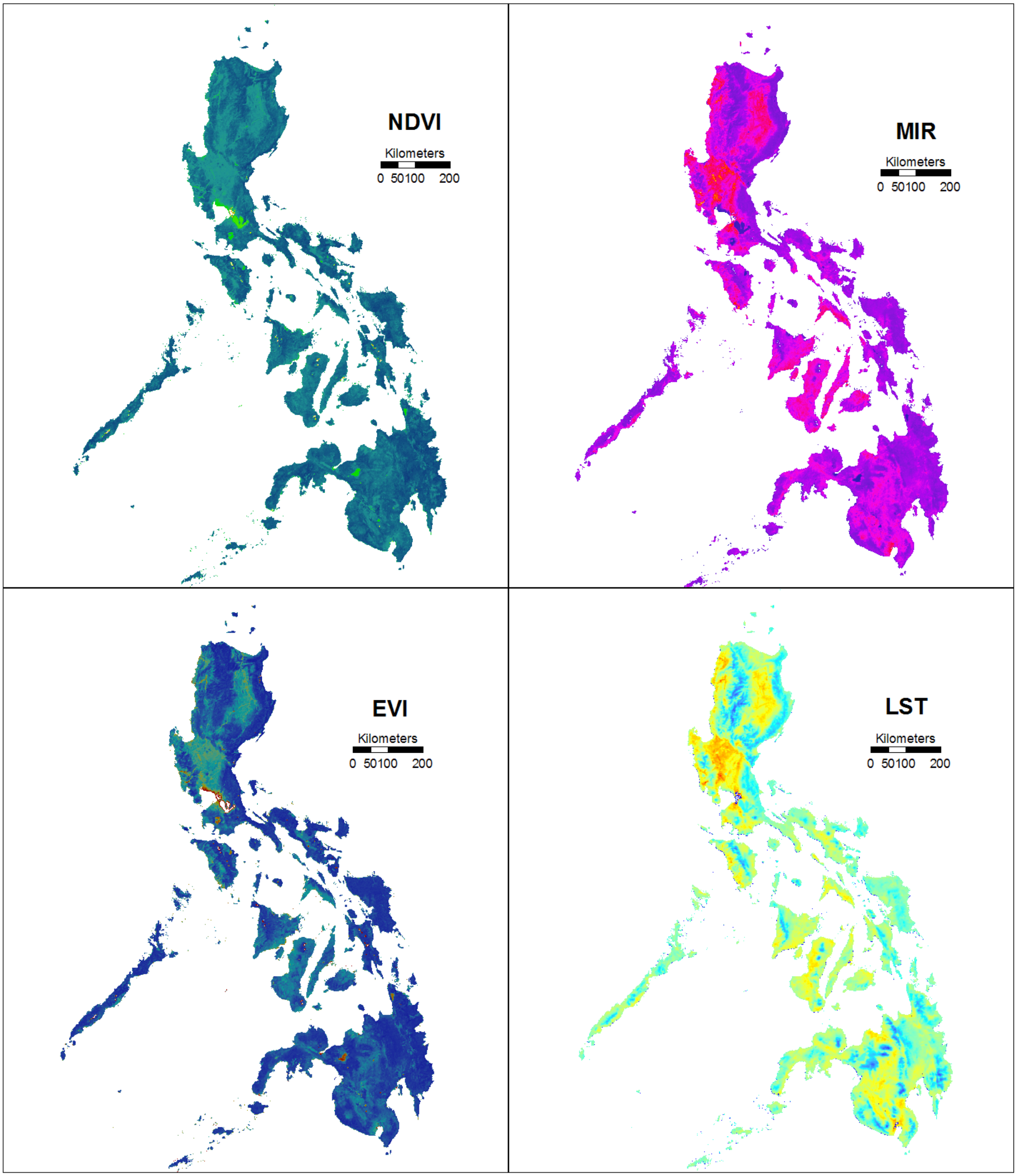
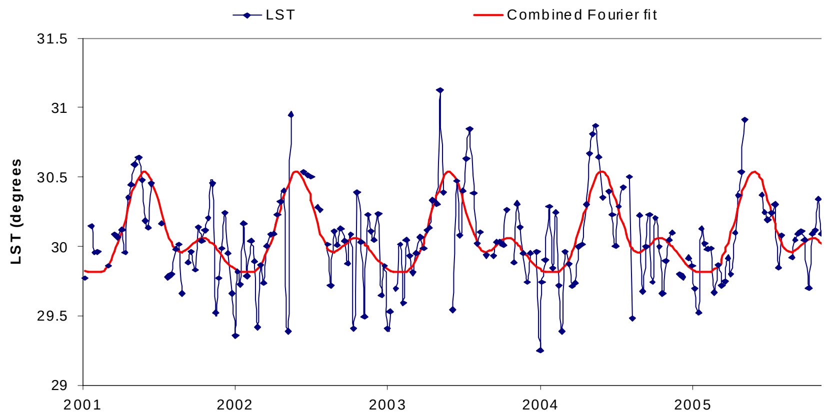


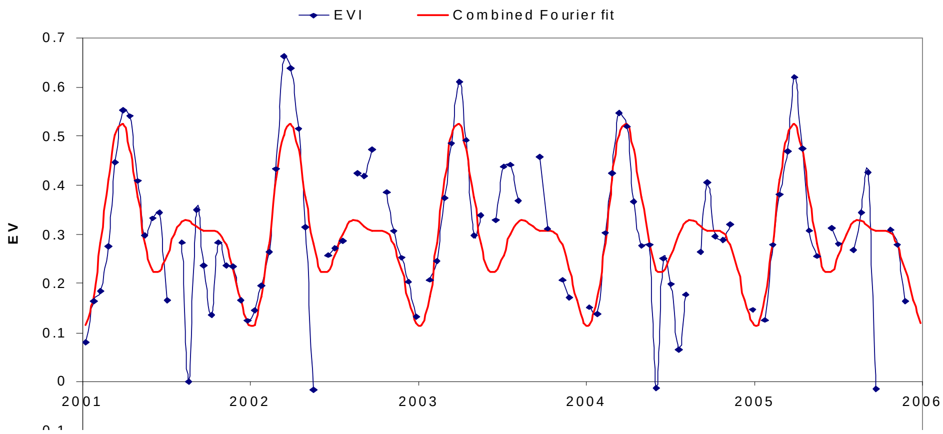
Figure 4: Time Series MODIS Parameter Values for Aparri, Philippines
MODIS values in blue; Fourier model fit in red

Location of Appari

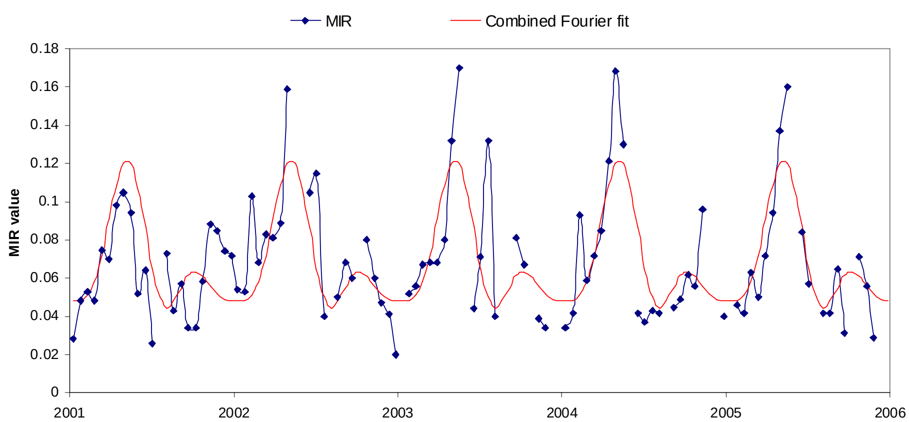
Land Surface Temperature



Enhanced Vegetation Index



Middle Infra Red



1.2 Temporal Fourier analysis and missing value interpolation

Time-series of satellite data (e.g. AVHRR, MODIS, etc.) provide multivariate datasets for each pixel. These time-series usually show strong serial correlation, *i.e.* repeated annual patterns, and therefore data redundancy. This may be eliminated by ordination methods, such as Principal Components Analysis or Temporal Fourier Analysis (Rogers, 2000). Temporal Fourier analysis (TFA) describes the cycles of temperature, vegetation indices, *etc*. as the sum of a series of sine curves with different amplitudes and phases.

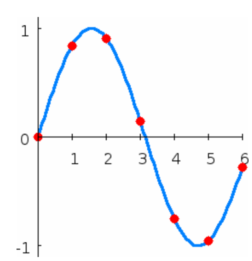
TFA removes data redundancy and produces orthogonal layers whilst retaining a description of seasonality. It also removes noise from the original satellite signal, as noise is generally at a higher frequency than seasonal events, so the corresponding harmonics may be omitted from the Fourier description to produce a smooth picture of seasonal change. Also, TFA achieves data ordination that has a biological interpretation in terms of cycles of seasonal events. This contrasts with and is preferential to alternative data reduction methods, such as Principal Components Analysis, where seasonal events may contribute in complex ways to any number of principal component axes and images derived from them (Rogers, 2000). Previously, TFA has been applied to long-term datasets derived at monthly intervals from the Advanced Very-High-Resolution Radiometers (AVHRR) on the National Oceanic and Atmospheric Administration (NOAA) series of satellites (Hay, 2006).

As described above, pre-processed and quality assessed MODIS datasets are provided at 8- or 16-day intervals. These intervals do not divide a whole number of times into a standard year of 365 days, which is a condition required for simple TFA to derive correct estimates of annual, bi-annual, tri-annual etc. cycles. Further, as MODIS datasets are available at relatively short intervals (8- or 16-day), compared to monthly maximum value composited AVHRR data, there are likely to be more missing values in the time-series due to cloud contamination, off-nadir observations or other sensor problems than in longer period data.

Figure 4 gives an idea of the number of missing or excluded values for a pixel near Apparri in the northern Philippines. The blue lines on the graphs plot values for MIR, LST and EVI and show a significant number of gaps in the records – usually due to cloud cover, or dust/haze.

This highlights the desirability of using a long time series of imagery in any analyses: single images, or even a series of imagery for a whole year may not be sufficiently complete to provide a reliable record. The Fourier processing can also be seen to smooth the outlying values, many of which are likely to be due to anomalous reflectance measurements due to adverse atmospheric conditions (cloud or dust), or sub-optimal viewing angles.

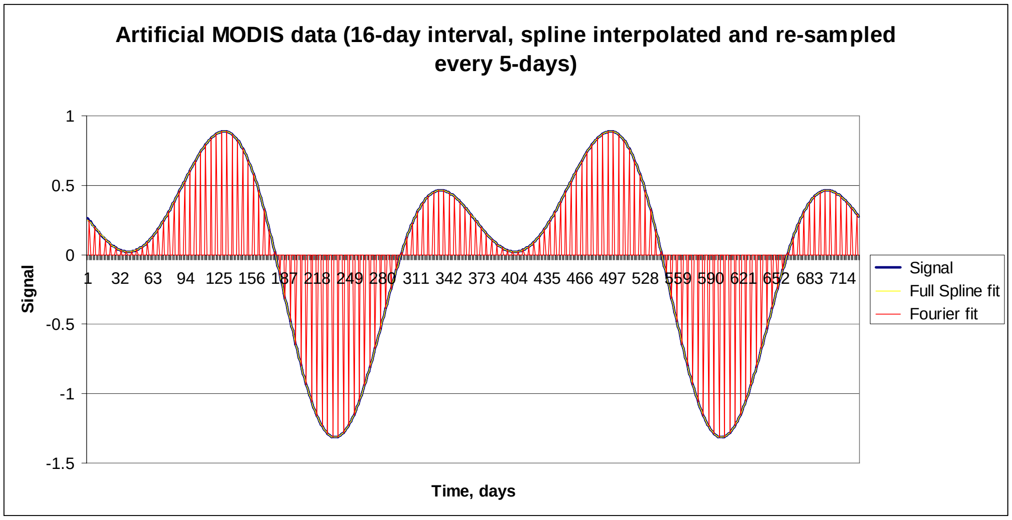
Figure 5: Splining



To overcome these two problems we interpolated the time-series using cubic spline interpolation before temporal Fourier processing. Temporal rather than spatial cubic spline interpolation was used, which fits a different cubic polynomial equation between each pair of known values to create a smoothed curve passing through all the known time series values (illustrated in Figure 5). If several consecutive values were missing, then the process was repeated until all missing values were replaced. Pixels with more than 80% of their values missing were discarded altogether, and set to the no data value.

Although daily interpolated time-series provided the most accurate estimates of Fourier amplitudes and phases for artificial test data, due to computational constraints a 5-day interpolated time-series (i.e. 73 values per year) was chosen that still provided accurate estimates of Fourier cycles. Figure 6 illustrates the procedure using synthetic data.

Figure 6: IMODIS Resampling and Spline Interpolation with Synthetic Data.



The red lines show the 5 day samples used for spline interpolation to create the yellow full spline fit, which almost perfectly matches the original signal.

The TFA algorithm was implemented in Microsoft QuickBASIC 4.0 and outputs 17 different TFA layers for each input product. The TFA layers include: the overall mean (a0); the amplitude of the annual (a1), bi-annual (a2) and tri-annual (a3) cycles; the phase (i.e. peak timing) of the annual (p1), bi-annual (p2) and tri-annual (p3) cycles in months (starting at zero in January); the proportion of variance in the original time-series that is explained by the annual (d1), bi-annual (d2), tri-annual (d3) and all three cycles combined (da); the minimum (mn) and maximum (mx) of the seasonal cycle recomposed from the first three harmonics only; and the variance (vr) of the original time-series. The relevance of these variables in summarised in at the end of this section.

In addition to TFA, the programme conducts and provides layers on quality control and error trapping procedures. The percentage of granules missing from the input time-series is recorded in a layer (e1) as well as the percentage of granules with values outside the permissible geophysical limits (e2), as specified in Table 1. Pixels with more than 80% missing granules, or granules with values outside the permitted limits were excluded from TFA. Further, the percentage of spline interpolated values that deviate by more than a specified amount (see Table 1) from the Fourier fitted time-series is recorded (e3). These values are then removed and interpolated from adjacent spline interpolated values. The temporal Fourier processing is repeated up to 20 times, or until no values deviate by more than the specified amount.

1.2.1 Relevance of the TFA Variables

The TFA output variables are well suited to distribution modelling as they reduce the huge range of possible data layers represented by a long and frequent time series (500–1.000 potential variables per climatic parameter, as shown in Table 1) to a manageable number of biologically meaningful layers. Figures 7 and 8 show four TFA variables for the Enhanced Vegetation Index.

The key feature to emphasise is that each parameter highlights different areas. The top maps of Figure 7 – proportion of first Component variance and total variance (d1, and vr) – show that vegetation index variability is greatest in to the north and west, but that this contributes little to the overall variance, in all but a few relatively small areas in the north.

Figure 7: Examples of TFA Variables for EVI

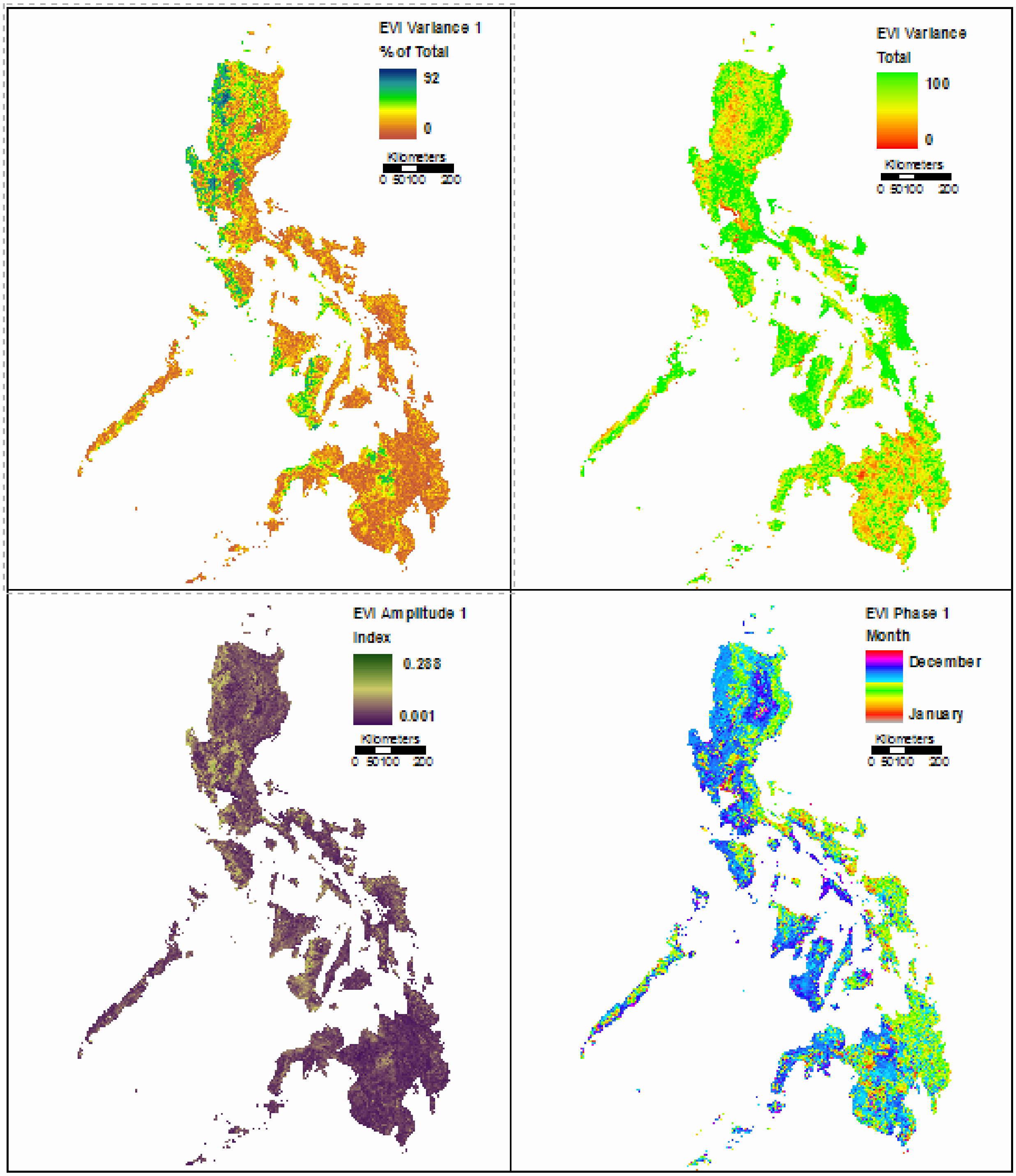
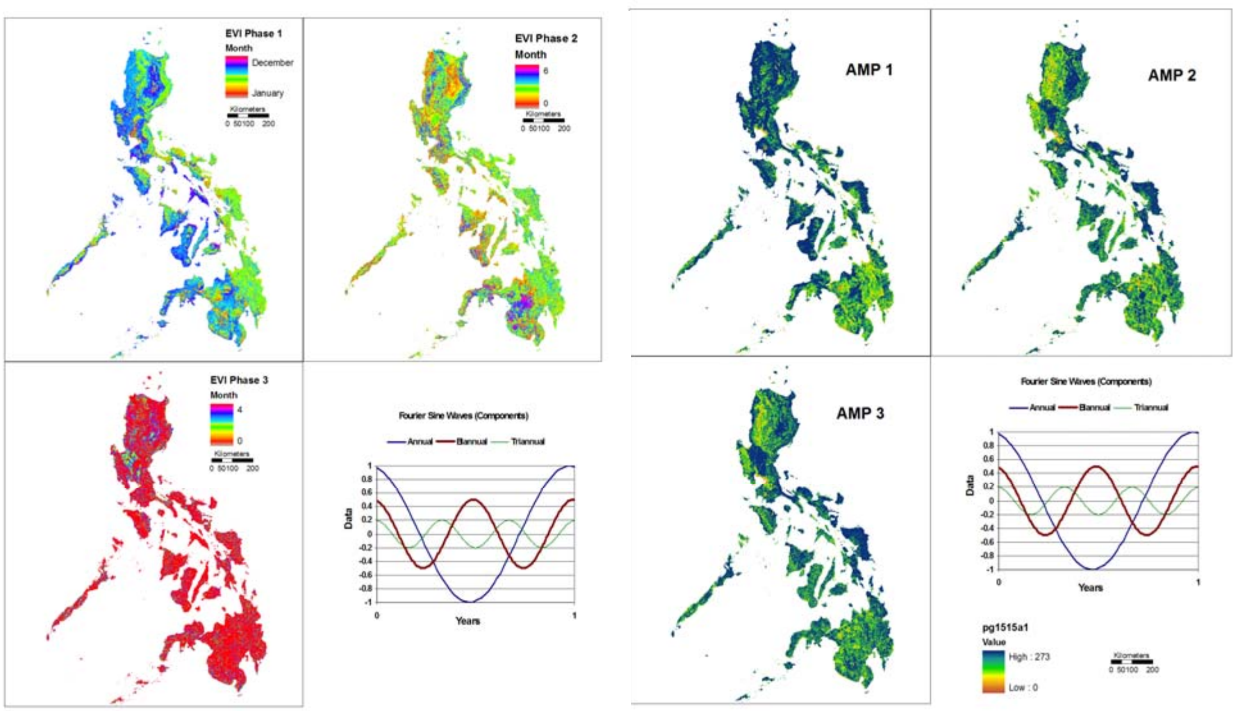


Figure 8, also for EVI, provides more detail of phase (p1, p2 and p3, to the left) and amplitude (a1, a2 and a3, to the right) for each component; again clearly illustrating that the different parameters highlight different parts of the country. The amplitudes, for example, showing the high of the respective sine curves, or the strengths of the annual, biannual and triannual cycles – are very different from each other. The annual cycle is weakest on the south, and particularly strong in the north-west, in a zone that is not picked out for the other two cycles. The phases also contrast substantially – with the biannual timing varying considerably across the country, unlike the tri-annual phase, which is relatively constant. As a consequence, each parameter is likely to add valuable and distinctive predictive power to the statistical modelling of animal distributions and disease risk.

Figure 8: Examples of Phase and Amplitude TFA variables for EVI



1.3 Post-processing

After TFA processing (described above) outputs for all products were mosaiced and georeferenced using a custom built program written in QBasic. The MODIS sinusoidal projection is restricted to MODIS products, which can make combination with other spatial data layers difficult. The TFA products were, therefore, reprojected to geographic projection using ArcGIS and Idrisi Andes, to facilitate further display and analyses, and to allow direct manipulation with other unprojected layers. The products were also converted to IDRIS 32, as well as ERDAS Imagine format, to facilitate further use.

Image Archive Description and Documentation

All data has been supplied on an external hard disk, which has been copied to BAI machines and portable hard disks. This includes the MODIS image archive, the data acquired by EAHMI, and data used for suitability mask generation. EAHMI will host a workshop specifically to address this issue in June 2007.

1.4 MODIS Image Archive

The TFA processed image archive supplied consists of more than 4,000 images. The files are provided in folder *\modis* of the external disk accompanying this report, together with a series of A2 format posters in a separate Annex. A catalogue of archive contents is given in Table 2.

Table 2: Image Archive Details.

Folder	Subfolder	Contents
\modis	\docsppt	Report files and PowerPoint presentations. *Faophilv03.doc* Step One report. *WintandtatemModisv06.doc* Step Two (this) report. *Faophilv1.ppt* Presentations given at seminar to BAI staff in Manila and training course in remote sensing and GIS applications at UPLB sponsored by EAHMI.
	\posters	GIS and map data for posters of MODIS data provided for BAI use. Posters can be displayed in ARCGIS *.mxd* files, with corresponding *pdf* printouts. Spatial data within subfolder as ArcGIS grids. Filenames for posters in full English, also provided in *.pdf* format for printing. PowerPoint *modis poster text.ppt* provides the text included in the ARGIS posters. Size A2, and can be changed by modifying the page size in ArcGIS. Includes MrSid format Landsat data, obtained from https://zulu.ssc.nasa.gov/mrsid/mrsid.pl, in a separate subfolder.
	\refs	Digital copies of most documents referenced in reports, as available.
	\windowed\lan	Original ERDAS *lan* format imagery from downloaded *hdf* files. Subfolders by parameter code (Table 3a, column EE) and then tile (Figure 2). File names as set out in Table 3B.
	\windowed\idrisitfa	TFA processed output in original sinusoidal projection (LST/BRDF) and geographic (Evapotranspiration). These are provided in case it is required to extend the area of the re-projected imagery to include the semi-permanent islands/reefs north of Palawan. Filenames as in Table 3A
	\TFAERDAS	TFA output in ERDAS imagine *img* format. Converted from Idrisi format using *modllrst2img.iml*. Will display directly in ArcGIS 9.X, but as automatically stretched images. Require classification before true max and min values displayed. Filenames as in Table 3A. Files organised in subfolders named according to parameter codes (Table 3a, column EE).
	\TFAIDRISI	TFA processed output in Idrisi format. Reprojected to geographic from sinusoidal, and masked using land water mask *phnoimsk.rst*. Idrisi macro. Filenames as in Table 3A. Files organised in subfolders named according to parameter codes (Table 3a, column EE).

Image file naming conventions are set out in Table 3A. The original imagery, converted from the original *.hdf* format, is provided for possible further use in ERDAS *.lan* format as indicated in Table 3B and Table 4. In all, there are just over 6,000 files, amounting to 10.7GB. Each has a statistics file associated with it. Spatial details of the LAN tile imagery are set out in Table 5.

Table 3:MODIS image files, Naming conventions.

A) Fourier Processed Files: ABCDEEPP.EXT.

A Location	B Projection	C Start Year	D End Year	EE Climate Parameter	FF Fourier Variable and Component	EXT File Type
P= Philippines	G=Geographic S=Modis Sinusoidal	0=2000	5=2005	03=MIR	A0=Mean	RST, RDC = Idrisi raster and document files
		1=2001		07=dLST	A1,2,3=Amplitude	IMG=ERDAS Imagine raster
				08=nLST	P1,2,3=Phase	
				14=NDVI	D1,2,3=Proportion total variance from each Component. DA=All	
				15=EVI	MN, MX, RN=Min, Max, Range	
				35 (ET)	VR=Variance of Time Series	
		V4=version4		90 (LW)*		

* LW=Land and water mask, MODIS version 4. Codes 1=Land 2=water, 3=ephemeral water

B) LAN Image Files: TTYYNNVV.EXT.

	YY Year	NN Image number	VV Climate Parameter	.EXT File Type
T6,t7,t8	2001	BRDF: 1-23	03=MIR	LAN=ERDAS7.4 Raster
u7,u8	2003	LST: 1-46	07=dLST	STA=ERDAS Statistics File
	2003		08=nLST	
See Figure 2	2004		14=NDVI	
And Table 4	2005	See Section 2	15=EVI	

The geographical metadata (projection, image boundary coordinates and so on) are held within the **.rdc* Idrisi documentation files. Geographic (latitude, longitude files) and sinusoidal files all have the WGS84 datum. The MODIS sinusoidal projection uses a spheroid of 6371007.181m.

Table 4: MODIS LAN Files: Tile Characteristics and Boundary Coordinates

ModisHVCode	t6	t7	t8	u7	u8
LongMin	117.0616000	111.7014000	110.0000000	121.8560000	120.0000000
LongMax	138.5578000	127.6973000	121.8517000	138.3388000	132.0056000
LongCen	127.6562856	119.5991568	115.8892935	129.9990835	125.9666234
LatMin	20.0000000	10.0000000	0.0000000	10.0000000	0.0000000
LatMax	30.0000000	20.0000000	10.0000000	20.0000000	10.0000000
LatCen	25.0679941	15.0390708	5.0127573	15.0390708	5.0127573
SINMinX	12231455.7160000	12231455.7160000	12231455.7160000	13343406.2360000	13343406.2360000
SINMaxX	13343406.2360000	13343406.2360000	13343406.2360000	14455356.7560000	14455356.7560000
SINMinY	2223901.0390000	1111950.5200000	0.0000000	1111950.5200000	0.0000000
SINMaxY	3335851.5590000	2223901.0390000	1111950.5200000	2223901.0390000	1111950.5200000
SINCenX	12787430.9935000	12787430.9935000	12787430.9935000	13899381.5190000	13899381.5190000
SINCenY	2779876.2644800	1667925.7717900	555975.2791100	1667925.7717900	555975.2791100
Shape_Leng	4447801.0532300	4447801.0532300	4447801.0532300	4447803.0193100	4447803.0193100
Shape_Area	1236433388070	1236433388070	1236433388070	1236434481160	1236434481160

The maximum and minimum values for each image are also held within the Idrisi image documentation files. The image values are scaled, so that the images can be stored as integer rather than real files, thereby saving space. Details of the rescaling criteria are given in Table 5.

Table 5: MODIS Image Values, Rescaling Criteria

Parameter	Fourier Variable	Image values are
MIR (03)	A0, A1, A2, A3, Min, Max, Var	Reflectance values * 10000
LST (07,08)	A0, A1, A2, A3, Min, Max, Var	(Degrees Centigrade+273)*50
NDVI (14) and EVI (15)	A0, A1, A2, A3, Min, Max,	Index Value * 1000
ET	A0, A1, A2, A3, Min, Max,	Mm/10days
NDVI (14) and EVI (15)	VAR	Value * 10000
ALL	D1,D2,D3,Da	Percentages
ALL	E1,E2,E3	Percentages
ALL	P1,P2.P3	Months*100. (Jan=1)

1.5 Additional data included on external hard disk

The additional data contained on the external hard disk are detailed in Table 6. These data represent the predictor data for the Philippines available at the time of writing (May 2007) and are envisaged as a core environmental animal health data archive, which will be expanded and refined as further data become available in the future. Much of the data has been assembled by EAHMI staff from various agencies and institutions, particularly the Bureau of Agricultural Research (BAR) and the National Statistical Office (NSO). Additional extracts of relevant global layers have been supplied, where an equivalent has yet to be found within the Philippines.

Table 6: Additional Data on External Hard Disk

Folder	Subfolder	Contents
\refs		Relevant references in PDF format.
\DATOS		Datakit of Official Philippine Statistics from National Statistics Office (NSO) through EAHMI.
\PHIGISBoundaryFiles		Administrative boundary files from Bureau of Agricultural Research (BAR) through EAHMI.
\Bai_eahi		Animal disease field record data from Bureau of Animal Industry (BAI) through EAHMI.
\Philag02		Philippines 2002 agricultural census data from National Statistics office (NSO) through EAHMI.
\Suit		Suitability masking data, used in UPLB training course. These replicate many of the layers provided elsewhere of the disk, and are collected in a single folder to facilitate access for teaching purposes.
	\faosuit	Data used to compile FAO global livestock suitability layers as detailed in the Word document within folder. ArcGISv8.3 format document file *Faosuit83.mxd* provided to display parameters supplied. Layer names file details can be obtained through the legends in this file .
	\philsuit	Data for the Philippines extracted from other folders on the hard disk for potential use in animal and disease suitability masking. Layers can be displayed in ArcGISv8.3 and above using *philsuit8.3.mxd*
\philcif		Philippine Countryside in Figures from Information Services, Publications & Archives Division (ISPAD) of the National Statistical Coordination Board (NSCB), ISSN 1656-0485, through EAHMI.
\globext		Extract of potentially relevant public domain global GIS layers: FAO Gridded livestock of the World (*faoglwmaps.mxd*); demography and water (*phgloextract.mxd*).

Future Requirements for Preparing Animal and Disease Modelling Datasets

The basic analytical process is broadly similar for all methods of distribution modeling, summarised in the Step One report (Wint and Tatem, 2007) and requires that all data are in, or are converted to, raster format image files for analysis. Modelling analysis initially involves establishing statistical relationships between the disease (either presence/absence or incidence/density) and a range of predictor variables including the imagery for a series of sample locations. These relationships are then applied to all locations (pixels) in the study area to provide a complete coverage predicted distribution.

1.6 Training Data and Suitability Masking

The first step is to select and prepare training data for the target variables – in the present context this means livestock densities and disease occurrences. As well as being the primary justification for the entire exercise, this choice will largely determine the variables identified as potential predictors for the modelling process. If there are a number of possible targets, then those with the most consistent and high resolution observations are likely to provide the best outputs. If possible, some index of occurrence should be used that controls for as many factors as possible (e.g. land area for polygon based data, and susceptible animal numbers), though if this is difficult there is always the fall back of simple presence and absence, which allows cumulative data of varying units to be converted into a single variable.

Suitability masking – using land identified as unsuitable for disease, vector and host animal to delineate areas where a disease or animal can be defined as absent – is a critical component of distribution modelling. Though there is likely to be considerable overlap, it is probable that different diseases and different animals will have different suitability masks, which should be defined separately.

It is important that the threshold values set for each component variable to define unsuitable land are conservative – to minimise the areas wrongly identified as unsuitable. Once defined, the unsuitability layers can be used in three ways: a) to mask the model outputs with know zeros, and help identify false positive predictions; b) to correct input training data that is in the form of densities per unit area within a series of polygons to be densities per square kilometer of suitable land; and c) to help select target variables that are less amenable to risk modelling because suitability masks cannot be satisfactorily defined for them.

It is recommended that suitability masks be prepared for all main livestock species: carabao. cattle, chickens, ducks, horses, small ruminants (goats and sheep) and pigs, with separate masks for small scale backyard and large-scale commercial production, although not all will be used for distribution modelling in the first instance.

1.7 Establishing a Set of Potential Predictor Variables

Another important step is to compile a standardised predictor archive from which values can be extracted for each analysis location. The final form of the extracted archive should be a ”flat file” of the form ”point id, point x, point y, disease status1, disease status 2, predictor 1, predictor 2, predictor 3, …. predictor x”, which means that the input data may be either point attribute values for the sample locations, polygon values extracted using a spatial join tool (assign data by location) or values extracted for each point from raster imagery. The image archive should thus ideally be a standard resolution and projection to simplify the extraction procedure, or at the very least collected into groups with a common resolution and projection.

The image archive can be derived from other raster images, or, in the case of polygon data, converted from polygon vector to raster imagery

It is also important to establish the precise sample points to be used for disease and predictor value extraction. These locations may one of a number of types: a) the disease monitoring sample points; b) a regularly spaced lattice of points extending over the area for which disease data are available; or c) the centre points of regions or polygons (such as administrative areas) for which disease data are available. Note that, providing the sample point frequency is comparatively high, it is usually preferable to extract point data from polygons, as otherwise the raster data must be averaged for each polygon, which will result in substantial loss of information. Further, only if there are multiple sample points per polygon will the statistical modelling process be able to take into account any variation in resolution between different predictors.

A number of methods are available to extract raster values: – Idrisi uses the EXTRACT module, which can produce ASCII tables of values for point, or mean, minimum, maximums and the like for polygons. ArcView uses the Summarize Zones on the Analysis tab; and ArcGIS uses the Extraction utilities in the Spatial Analyst Tools. Various standalone custom written point value extraction tools in Visual Basic and R are also available.

1.7.1 Candidate Variables for a Predictor Archive

The central subject of this document – MODIS satellite imagery – form the core of a disease modeling predictor archive – not only because of their inherent properties as climatic indicators with links to temperature and vegetation, but also because of the seasonality that derives from the TFA processing. Other indicators are, however, also likely to be useful as predictors – including demographic, topographic, hydrological and agricultural variables. These layers may also contribute to the suitability masks needed to add to the known zero values, and enhance predictive accuracy.

The potential predictors currently available within the EAHMI archive have been detailed in the report accompanying Step One of this consultancy, but are summarised in Table 7 for convenience, together with a note of the additional processing currently needed. In general, categorical variables are less useful as predictors, unless they are used to delineate zones into which the modelling can be broken down. They are, however important inputs for suitability masking and so are listed here separately.

All the predictors require converting to rasters, usually at 1km resolution to match the remotely sensed environmental parameter. The elevation and population related layers are, however, too spatially detailed for this resolution and should be stored as 100m resolution rasters to preserve as much information as possible.

Some additional processing is needed for selected layers: the provincial agricultural layers will require suitability correction if they are to be used as predictors – most especially the livestock population data and the extent of cultivation layers; and the infrastructure (roads, built-up areas, rivers and lakes) should be converted to distance to feature layers, thereby adding these variables to the continuous value predictor suite.

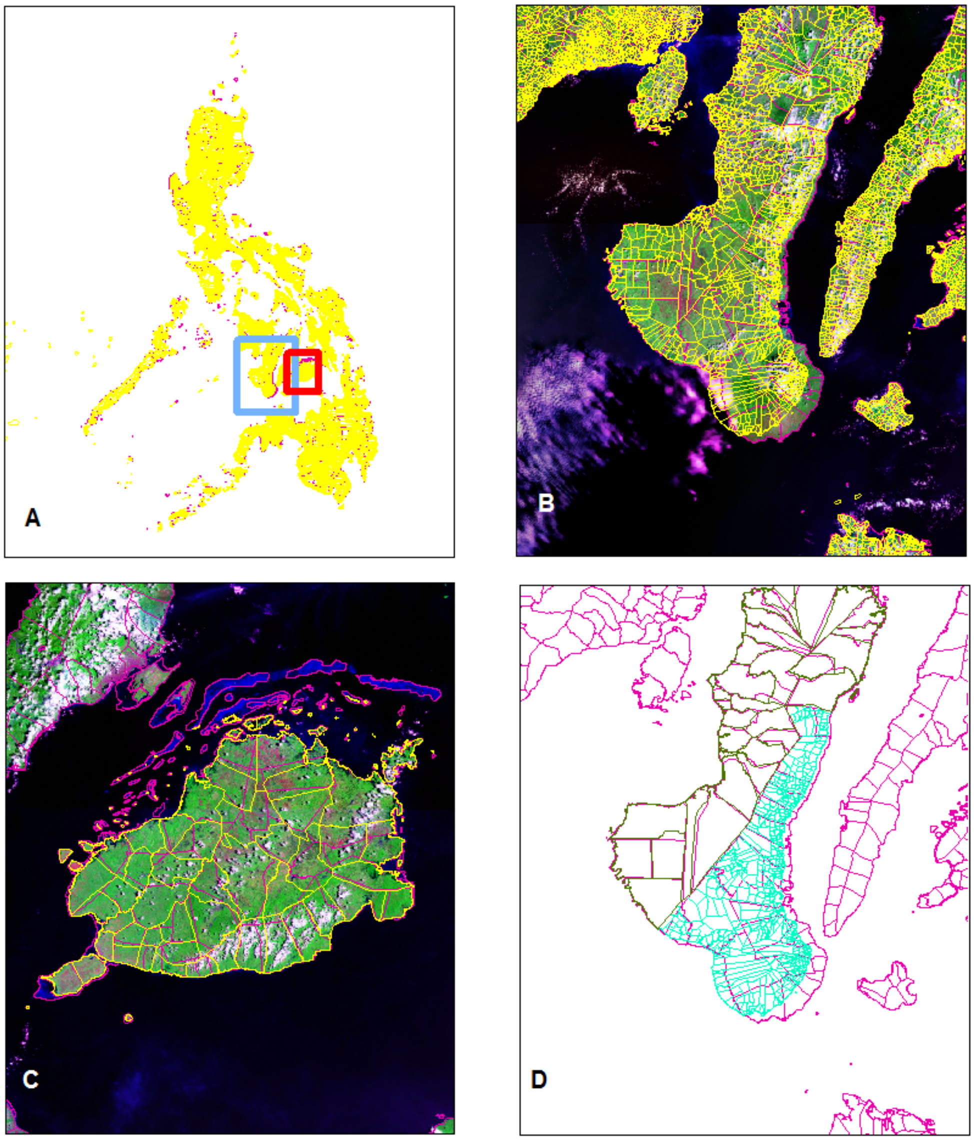
Table 7: Candidates for Predictors and Suitability masking Criteria

Category	**Variable**	**Type**	**Resolution**	**Further Processing Required**
Continuous				
Remotely sensed	nLST, dLST, MIR	Raster	1km	None
TFA	NDVI, EVI	Raster	1km	None
	ET	Raster	1km	None
DEM	Elevation, Slope	Raster	100m	None
Land Cover	Global Tree Cover	Raster	1km	None
Demographic	Global GPW3	Raster	1km	None
	NSO Population	Vector	Barangay	Adjust boundaries and rasterise
	NSO socio-economic	Vector	Barangay	Adjust boundaries and rasterise
	Poverty	Vector	Municipality	Rasterise incidence Derive number of poor people per square kilometer
Agricultural	All Livestock, Census 2002 -2006	Vector	Province	Rasterise and suitability mask
	Poultry	Vector	Barangay	Adjust boundaries, rasterise and suitability mask
	Cultivated Area	Vector	Province	Rasterise and suitability mask
	Crop Production Rice, Corn, Bananas, Mango, Coconut, 2002	Vector	Province	Rasterise and suitability mask
Infrastructure	Roads, Railways, Built up areas	Vector		Convert to distance to feature rasters
Water	Rivers, Lakes	Vector		Convert to single distance to open water rasters
Categorical				
	Ecozone	Vector		Rasterise if used
	Forest	Vector		Rasterise if used
	Land Cover	Vector		Rasterise if used
	Soils	Vector		Rasterise if used
	Geology	Vector		Rasterise if used
	Rice categories	Vector		Rasterise if used
	River Catchments	Vector		Fill missing portion and rasterise

Though the data collated at EAHMI is of generally very high quality, and is largely mutually compatible, the barangay level human population data will need some manipulation to ensure its usability, as it will be important to have these values at a better resolution than provincial in order to be an effective predictor. This is because in some places the barangay basemap is distorted relative to the BAR basemap, and thus needs to be manipulated to ensure the two sources match.

The discrepancies between the two data sources are illustrated in Figure 9, in which the yellow lines are the NSO barangay population data boundaries, with purple BAR municipality maps underneath. The two coastlines do not match where the purple is visible, as shown in Fig 9b in which the BAR municipality base map (purple) is overlain on a Landsat image, which matches well, thereby validating the BAR map coastlines. The barangay maps are made up from a number of sub-regional or provincial files (turquoise boundaries in Figure 9d), only some of which appear to need major rectification to get the coastlines of the two boundary files to match. There is also a wider problem that there are mismatches in municipality boundaries (Figure 9c).

Figure 9: Administrative Boundary Mismatches.



Given that the primary requirement is to transfer the population information to a reliable basemap, two options are open: a) to rectify the necessary barangay sub-files, and recompile a national barangay boundary map; and b) to rectify the boundaries, and assign the density data to the existing BAR municipalities - possibly by rasterising the corrected barangay map to a 100m resolution density layer, and then extracting mean values for each BAR municipality. Checks would then needed to make sure that numbers from this process produced populations that were close to the official figures.

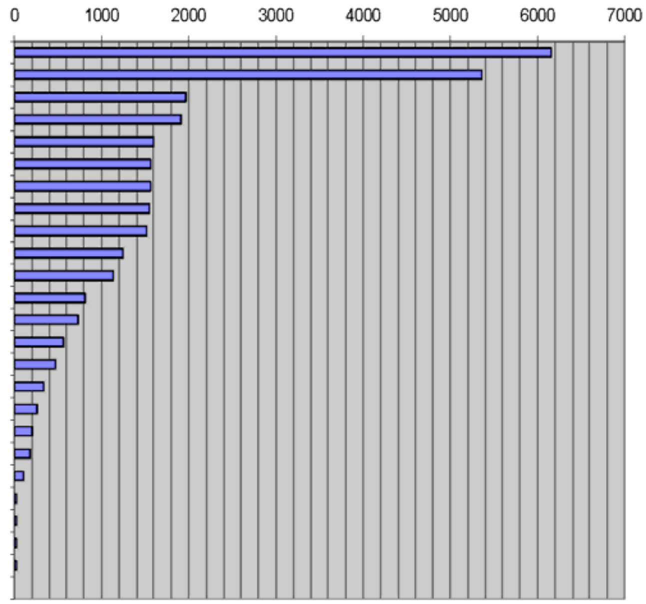
A third option may also be available using either a) a partially completed BAR barangay boundary map to replace the NSO boundaries in the relevant sub-files, providing, of course the BAR map covers the appropriate areas; or b) a barangay boundary map produced by University of California at Berkeley and the International Rice Research Institute at Los Baños (Hijmans *et al*, 2007)). Checks would have to be made to ensure municipality boundaries matched the BAR basemap.

1.8 Animal Diseases

A substantial amount of animal disease data has been assembled and cleaned by BAI ICTU and EAHMI (Catbagan *et al,* 2007) and is held within the EAHMI archive. The great majority are aggregated and geo-referenced to province level. The number of records for each disease is summarised in Figure 10 below, which range from the very widely reported fasciolosis and hemorrhagic septicemia to very rarely reported diseases, such as coccidiosis and brucellosis.

Figure 10: Disease Records for 1997-2004, as Compiled by EAHMI.
Source: Catbagan *et al.* (2007).

Disease	Total
Fasciolosis	6,160
Hem_Sept	5,357
H_Cholera	1,968
S_Flu	1,909
CRD	1,596
Coryza	1,565
S_Plague	1,563
A_Pox	1,547
NCD	1,514
Surra	1,249
S_Erysipelas	1,137
F_Cholera	805
A_Coccidiosis	733
Orf	558
Eph_Fever	468
FMD	342
A_Malaria	258
Colibacillosis	207
Blackleg	182
Anthrax	97
S_Pox	24
Tetanus	21
S_Dysentery	18
Coccidiosis	17
Brucellosis	6
Grand Total	**29,301**



There is thus a wide range of potential targets for provincial disease distribution modeling. Discussions with EAHMI and collaborators to select the priority target diseases centered on a number of issues, revolving primarily about: the availability of reliable and geo-referenced sub-provincial level information that could be used to refine the spatially coarse provincial data; the likelihood that the disease distributions could be related to environmental parameters, and the diversity of supplementary information (markets, slaughterhouses, animal movements and so on), that might be used to enhance the basic disease data.

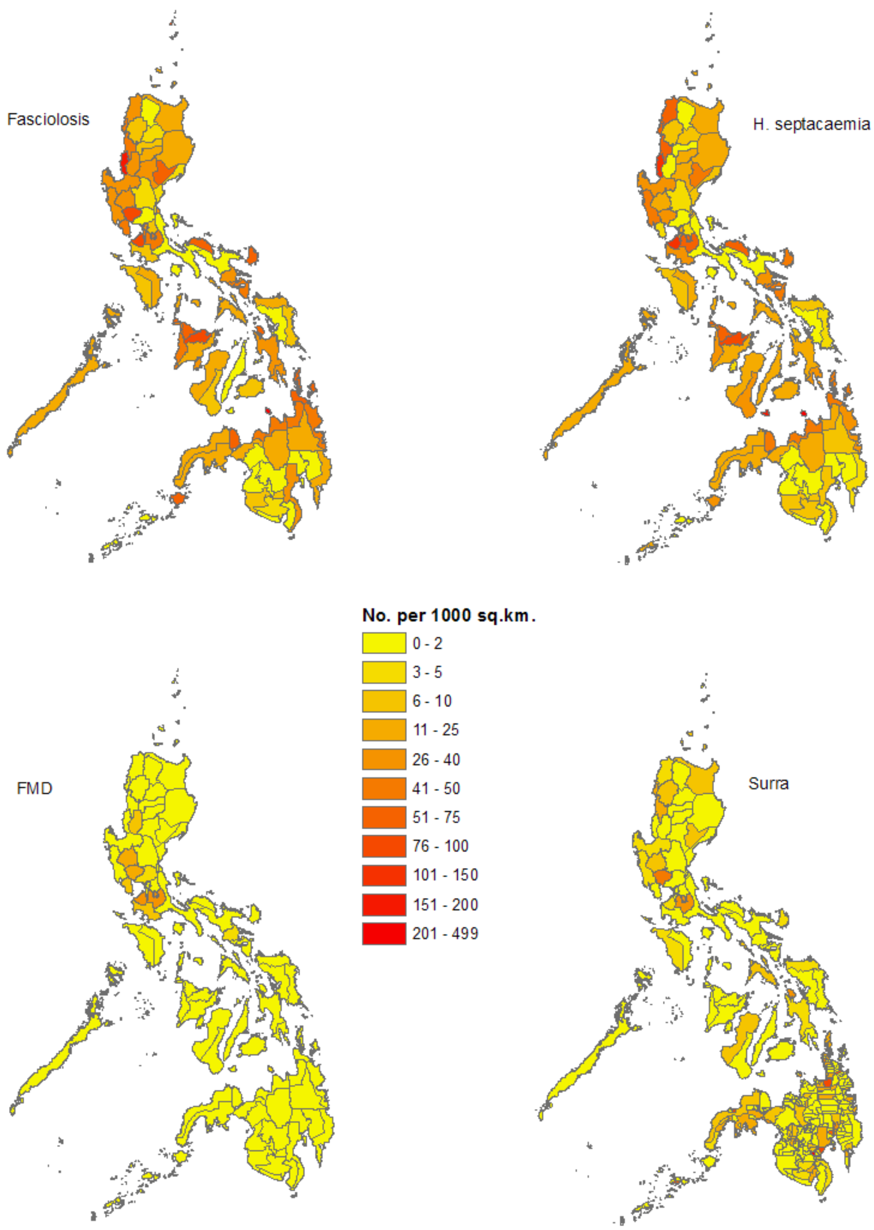
The two most commonly recorded diseases – fasciolosis and hemorrhagic septicemia - were identified as important target diseases by dint of their frequent occurrence. The former's transmission by snails points to a strong link with environmental conditions, though it was agreed that the latter might not be such a suitable target because the mechanisms through which environmental factors might affect the disease were less clear.

Two other diseases – surra and FMD – present themselves as potential candidates for further investigation and risk mapping. Surra is a widespread, vector borne disease of rural areas, which has been subject to extensive surveillance by the Mindanao Unified Surra Control Approach (MUSCA) project. Philippines has recently been declared free of FMD, but active surveillance continues and interest has been expressed in a risk assessment based on the environmental characteristics of previous outbreak locations for better targeting of field operations. Substantial datasets exist for both diseases, which makes them particularly amenable to risk modelling. The merging of provincial and barangay data will require careful scrutiny to avoid duplication and ensure that any remaining provincial values are adjusted to take account of the barangay data.

Except for surra in Mindanao and FMD in Luzon, all currently available disease records are aggregated to provincial level and cannot yet be disaggregated. Wherever barangay or municipal level disease data are available, such as for surra and FMD, they should be incorporated into national datasets because of their finer resolution. Provincial data can thus be upgraded piecemeal according to availability, and providing the compensation of provincial data is carried out correctly. The integration of barangay data into the provincial data need not therefore use all the barangay values, should spatial matching be incomplete.

The provincial distributions of fasciolosis, hemorrhagic septicemia, surra and FMD are shown in Figure 11, as total occurrences per 1,000 square kilometers.

Figure 11: Four Possible Target Diseases



Most of the disease data currently held by EAHMI relates to reported field records of suspected disease outbreaks and runs to the end of 2004. If more recent information exists, it should be appended to the existing dataset of field records for completeness. There is also finer resolution laboratory based information, which would be especially valuable to acquire as it could be used to corroborate the field disease occurrence records.

There are several possibilities for the final format of the disease training data, depending on the degree to which combinations and amalgamations of the various scales of data are successful. At the least, values can be converted to presence and absence, which can then be modeled using logistic regression based techniques. It would be preferable, however, to have some continuous variable such density (cases/sq. km.) or some index of incidence (*e.g.* cases per susceptible animal) as these units are likely to provide a better basis for risk modelling. Care will have to be taken to distinguish between measures of sero-prevalence (which may reflect vaccination as much as disease history), as indicated by the CATT data for surra, and other measures more closely related to actual disease occurrence.

1.9 Suitability Mapping

Discussions with stakeholders in Manila and trainees at the RS and GIS training course in UPLB have confirmed that ample capabilities exist within EAHMI collaborating institutions to construct feasible suitability maps. The Los Baños training course should have enhanced these skills further, and enable EAHMI to produce suitability masks for all selected livestock species and diseases for modelling. A training course to further improve national skills in these techniques is planned for October 2007.

It is strongly recommended that suitability thresholds for the selected diseases, and indeed the livestock species, are defined at he same time, and for the whole country rather than for specific parts of the Philippines alone, comparable in principal to the global livestock suitability definitions used by FAO described in the first report of this series. Similarly the masks produced should be for year round occupancy, rather than seasonal minima or maxima, as the distribution modelling will, at least in the first instance be addressing annual data, rather than intra-annual variation. This will most probably entail discussions between all interested parties, including IT specialists, epidemiologists and veterinarians, so that the many overlaps between diseases and between animal species can be rationalized effectively

The choice of criteria by which to define suitability will depend not only on the target variables, but on the parameters available within the EAHMI archive. Examples of the sort of agro-climatic parameters used by FAO for its global suitability product have been set out in the first report of this series, and a number of others, including wetlands, distance to built-up areas, water or roads or markets, were discussed within EAHMI as being potentially useful.

It may well be prudent to produce suitability masks for both intensive and backyard systems, as they may well differ substantially. It is, for example, possible that intensively reared animals are less likely to be found in areas with low human population density, far from markets and built-up areas, than are back yard animals. It is also possible that there is a degree of similarity between the suitability thresholds for all intensively reared livestock species, as the influence of anthropogenic factors is likely to outweigh the environmental ones.

Recommendations for Step Three of Consultancy

A major priority for EAHMI and partners over the next six months is to ensure that a series of multivariate analyses are undertaken to determine the environmental signatures of animal populations and selected diseases, so that a set of disease risk maps can be prepared for internal scrutiny and discussion within BAI and consideration of EAHMI's Tri-Partite Review (TPR), scheduled for late November/early December 2007 (Bourn, pers. comm.). Achieving this objective will involve the three stages set out in Section 5:

a) Selection of diseases and preparation of training data;

b) Selection and cleaning of predictor archive data; and

c) Multivariate analysis, modeling and mapping.

In the first instance and as proof of principal, it is recommended that the following animal diseases and livestock species be targeted for disease risk mapping:

- Surra (*Trypanosoma evansi*), associated with carabao, cattle, horses and pigs;
- Fasciolosis (liver-fluke), associated with carabao and cattle;
- Hemorrhagic septicemia, associated with carabao and cattle; and
- Foot and Mouth Disease, associated primarily with pigs in the Philippines.

The very substantial range of spatial data available from BAR, in particular, and the Philippines in general, has been summarized in the previous section. There is also capacity to manipulate and process these data to a high standard, and though the number of staff is somewhat limited and dispersed amongst a number of units and institutions (e.g. BAI, BAR and UPLB), close collaboration between these groups should ensure access to the necessary skills. Significant remote sensing expertise is also available in the Philippines (e.g. NAMRIA and UPLB) and has been widely used in the country, though not yet in the field of animal health and production.

There is, therefore, considerable national potential for expanding the largely vector based skills within the EAHMI collaborator network to include the basic raster manipulation and production processes needed to finalize the production of "clean" disease training data and predictor archive that underpins the distribution modeling process.

It is, therefore, strongly recommended that as much preparatory work as possible is carried out in the Philippines to minimize the need for costly external involvement in the preliminary processing – including data cleaning, formatting and manipulation; suitability masking for both animals and diseases; and construction of a predictor archive of standardised rasters matching the bounds and resolution of the environmental imagery delivered to EAHMI by this consultancy.

The local implementation of these steps has several advantages beyond limiting cost, the most important of which is that the people involved with the data preparation and suitability masking know the country, the diseases, and the databases, far better than most external analysts, and are therefore more able to make realistic judgments. Local knowledge is particularly valuable in the identification of suitability masks, and input from several collaborating sources should ensure their reliability and, equally importantly, the acceptability of the definitions to national users.

External input would then only be required for the interactive modelling and reporting phases of the analyses, to be undertaken over an estimated 12-14 week period from receipt of the input data layers, and involve approximately 3 days input per target variable selected for modelling. Additional resources would be required for reporting, and, if desired, delivery dissemination and training in the Philippines. A provisional schedule of activities is shown in Figure 12 and estimates of external professional input requirements in Table 8.

Table 8: External Resources for Step Three: Multivariate Modelling

Activity	Days Required
Extraction of data from prepared databases	2
Distribution modelling of four target livestock species: carabao, cattle, pigs and horses	12
Disease distribution modelling of 4 target diseases: surra (*Trypanosoma evansi*), fasciolosis (liver-fluke), hemorrhagic septicemia and foot and mouth disease	12
Reporting	7
Delivery and dissemination visit to Philippines (five working days in country, plus two travelling	7
Total	40

Figure 12: Schedule Envisaged for Step Three Multivariate Modelling

Activity	Apr	May	June	July	Aug	Sept	Oct
Submission of draft Step Two recommendations by ERGO	x						
Review and revision of draft Step Two recommendations by EAHMI		xxxx					
Data preparation by EAHMI	x	xxxx	xxxx				
Completion of transfer of documented disease training and predictor data to ERGO by EAHMI			x				
Data extraction for modeling by ERGO**				xx			
Risk modeling by ERGO**				xx	xxxx	xxxx	
Delivery of draft report on Step Three risk model outputs to EAHMI by ERGO**							x
Dissemination of model outputs and further training in Philippines* by ERGO**							xx

X = one week. * It is envisaged that this activity should coincide with the second part of RS and GIS training at UPLB, provisionally scheduled for 15-26 October 2007.
Timiing of ERGO activities dependent on timely delivery of required data by EAHMI

The following tangible outputs of step three of the consultancy are therefore envisaged:

- Technical report reviewing entire disease risk mapping process;
- Digital disease risk maps for GIS applications;
- PowerPoint presentation of risk mapping process;
- Poster illustration of risk mapping process;
- Capacity building through group discussions, presentation of seminar to BAI staff and participation in RS and GIS training course at UPLB;
- Updated data archive delivered to BAI ICTU and staff familiarized with its contents.

References

Boyd, D.S. & Petitcolin, F, 2004. Remote sensing of the terrestrial environment using middle infrared radiation, International Journal of Remote Sensing, 25, 3343-3368.

Cleugh, H.A., Leuning, R., Mu, Q. and Running, S.W., 2007. Regional evapotranspiration estimates from flux tower and MODIS satellite data, *Remote Sensing of Environment*, **106**, 285-304**.**

Catbagan, D., Bourn, D. Crescencio, R., and Molina, J., 2007. Preliminary Assessment of the Frequency and Distribution of Animal Disease Field Records in the Philippines: 1997-2004. Drfat produced for Comment by EAHMI, BAI, Manila, Philippines.

ERDAS, 2006. ERDAS field guide. ERDAS Inc, Atlanta, USA. http://support.**erdas**.com/documentation/files/**FieldGuide**.pdf Accessed 31 May 2007

Hay, S.I., Tatem, A.J., Graham, A.J., Goetz, S.J. and Rogers, D.J., 2006. Global environmental data for mapping infectious disease distribution, Advances in Parasitology, 62, 37-77.

Heute, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric performance of the MODIS vegetation indices. Remote Sensing Environ. 83, 195–213.

Hijmans, R., Garcia, N., Kapoor, J., Rala, A., Maunahan, A., and Wiecszorek, J. (2007) Global Administrative Areas version 0.6. http://biogeo.berkeley.edu/gadm/. Accessed May 31 2007

IDRISI Andes 2006. Image processing software. http://www.clarklabs.org. Accessed May 31 2007

Mu, Q., Heinsche, F.A., Zhao, M. and Running, S.W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, *Remote Sensing of Environment*, **in press.**

Rogers, D. J. 2000 Satellites, space, time and the African Trypanosomiases. In *Advances in Parasitology*, vol. 47 (ed. S. I. Hay, S. E. Randolph & D. J. Rogers), pp. 128-171. London: Academic Press.

Salomon, J., C. B. Schaaf, A. H. Strahler, F. Gao, Y. Jin, 2007. Validation of the MODIS Bidirectional Reflectance Distribution Function and Albedo Retrievals Using Combined Observations from the Aqua and Terra Platforms, IEEE Trans. Geosci. Remote Sens., in press.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N., Xiaoyang, Z., Jin, Y., Muller, J.-P., Lewis,P., Barnsley, M. J., Hobson, P. H., Disney, M. I., Roberts, G., Dunderdale, M., Doll., C., D'Entremont, R. P., Hu, B. and Liang, S., Privette, J. L. and Roy, D., 2002. First operational BBRDF, albedo nadir reflectance products from MODIS, Rem. Sens. Environ., 83:135-148.

Wan, Z., Zhang, Y., Zhang, Q. & Li, Z., 2002. Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data, Remote Sensing of Environment, 83, 163-180

Wint, W, Tatem, A., 2007. Remote sensing options and image processing requirements for spatial analysis of animal disease distribution in the Philippines: step one report prepared by Environmental Research Group Oxford Limited for the Environmental Animal Health Initiative, FAO, Manila, Philippines.

Appendix 1: Terms of Reference

The TOR applicable to the current consultancy are indicated below in ***bold italic***.

Terms of Reference for an International Consultant to Review Remote Sensing Options and Image Processing Requirements for Spatial Analysis of Animal Disease Distribution in the Philippines

As indicated in the project document and inception report, the Philippine Environmental Animal Health Management Initiative (EAHMI) (GCP/PHI/050/ITA) requires a sequential series of highly specialized, international consultancy inputs to:

1. Review remote sensing options and image processing requirements, and identify the most appropriate form(s) of remotely sensed imagery for spatial analysis of animal disease distribution and risk modelling in the Philippines;
2. ***Acquire and process selected imagery, prepare a paper and poster on how the imagery was obtained and processed, and its potential uses for presentation at a remote sensing and geographical information system training workshop in the Philippines;***
3. Multivariate spatial analysis of animal disease distribution and risk modelling; and, subject to obtaining meaningful results from the foregoing,
4. Model the potential spread of selected diseases under various scenarios.

Such specialized skills and expertise are not currently available in the Philippines.

Terms of reference for the first two steps of the above sequence are as follows:

Step One (Completed January 2007)
Conduct a desk study of remote sensing options and image processing requirements for the determination of land cover, climatic, seasonality and other environmental variables suitable for spatial analysis of the Philippines, with particular reference to high frequency, moderate resolution imagery, such as MODIS and others. The desk study should include:

- Review of recent publications and study reports relating to spatial analysis of livestock resource distributions and animal disease risk assessments commissioned by FAO AGAH and other agencies;
- Review and summarize the attributes of various forms of remote sensing, including: variables that can be derived, resolution, frequency, indicative costs and lead times;
- Identify/recommend the most appropriate form(s) of remotely sensed data for spatial/ environmental analysis of disease distribution and risk assessment in the Philippines;
- Provide estimates of image acquisition and processing costs;
- Prepare a concise report of findings and recommendations for the next step in the sequence of consultancy inputs, to be submitted by the end of January 2007

Step Two
Subject to review of recommendations and indicative costs identified in step one, it is anticipated that the consultant would subsequently be commissioned to:

- ***Acquire and process the recommended imagery in a form suitable for multivariate analysis and for use as thematic data layers in ArcView 9.1/2;***
- ***Prepare a paper and poster on how the imagery was obtained and its potential uses;***
- ***Undertake a two-week GIS/RS capacity building mission to the Philippines;***
- ***Install/transfer imagery/databases to EAMHI hardware at the Bureau of Animal Industry, Quezon City, Metro Manila;***
- ***Familiarize EAHMI staff and partners with imagery/database;***
- ***Present paper on remote sensing and spatial analysis in disease risk assessment at a seminar/workshop to be organized by EAHMI and the Institute of Renewable Natural Resources, College of Forestry and Natural Resources, University of the Philippines at Los Banos, planned for late March 2007;***
- ***Review status of EAHMI's animal disease and environmental database, identify gaps and discuss potential enhancements;***
- ***Prepare a comprehensive report of work undertaken, findings and recommendations for steps three and four, by the end of April 2007.***

Time Frame

Step One to be completed by end of January 2007

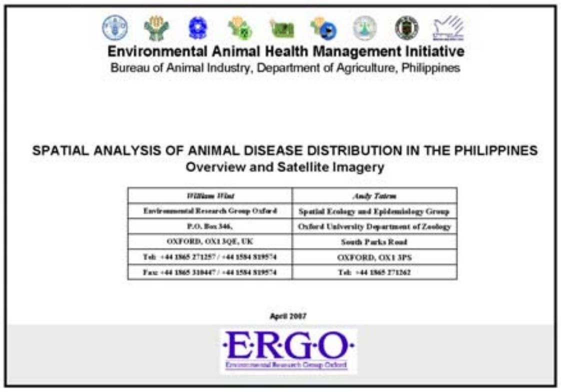
Step Two to be completed by end of April 2007

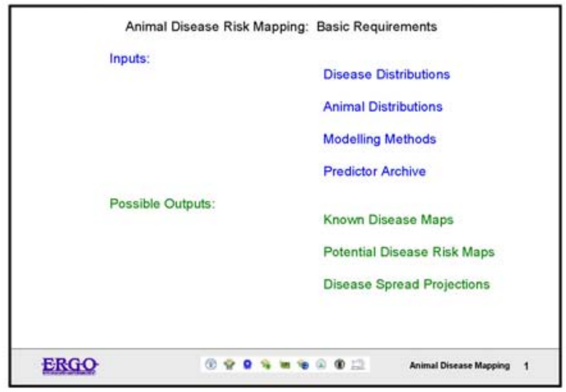
Appendix 2: Itinerary

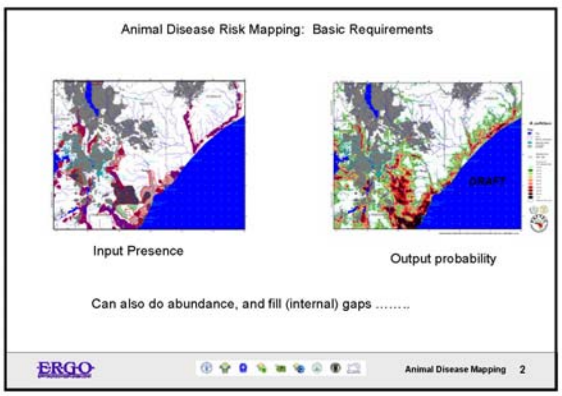
Sunday April 8	Depart London	
Monday April 9	Arrive Manila	Contacts
Tuesday April 10	Discussions with EAHMI staff and meeting ICTU, IT, BAI Director and FAOR	BAI Director Catbagan EAHMI staff: Bourn; Molina BAI Staff: Cresenscio, Morales, P and M Gealone and others
Wednesday April 11	Install data at BAI server, back up data. Meeting BAR BAR spatial analysis unit	Gealone, Abunda
Thursday April 12	Two seminar presentations	30+ BAI staff
Friday April 13	Meetings ICTU	Morales, Gealone and other ICTU staff
Saturday April 14	Assess data available	
Sunday April 15	--	
Monday April 16	Training	Bantayan, presenters, trainees
Tuesday April 17	Training UPLB, Seminars	10 participants
Wednesday April 18	Assess data	
Thursday April 19	Assess data. Wrap up meeting with Director BAI	
Friday April 20	Familiarize BAI ICTU with expanded database. Depart Manila	
Saturday April 21	Arrive London	

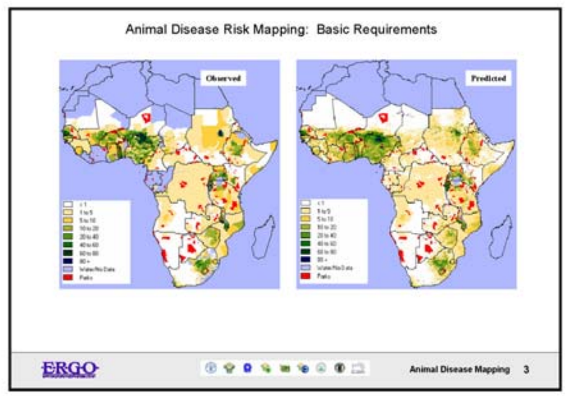
Appendix 3: Seminar/Training Course Presentations

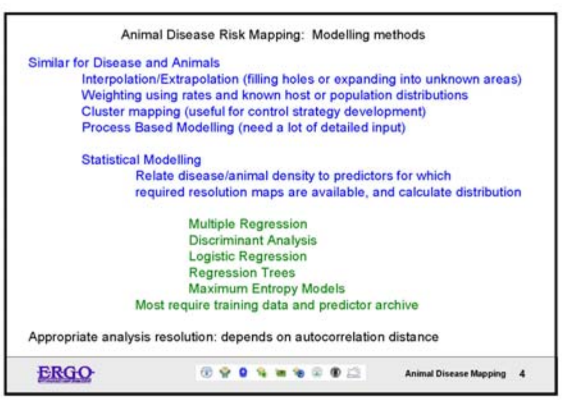
The following pages contain the handouts for the presentation given to BAI and collaborators in Manila and at the training course held at ULB in Los Bañyos

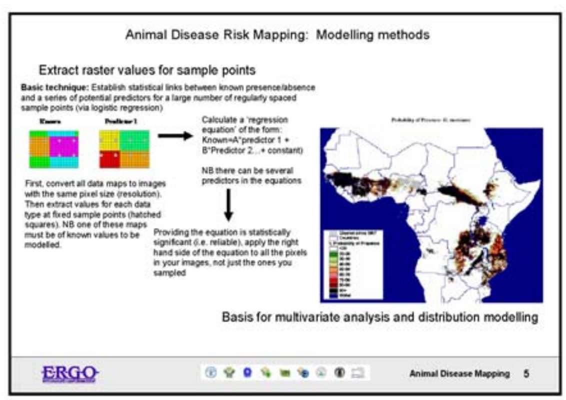

Environmental Animal Health Management Initiative
Bureau of Animal Industry, Department of Agriculture, Philippines
SPATIAL ANALYSIS OF ANIMAL DISEASE DISTRIBUTION IN THE PHILIPPINES
Overview and Satellite Imagery
William Wint
Andy Tatem
Environmental Research Group Oxford
Spatial Ecology and Epidemiology Group
P.O. Box 346,
Oxford University Department of Zoology
OXFORD, OX1 3QE, UK
South Parks Road
Tel: +44 1865 271257 / +44 1584 819574
OXFORD, OX1 3PS
Fax: +44 1865 310447 / +44 1584 819574
Tel: +44 1865 271262
April 2007
ERGO
Environmental Research Group Oxford

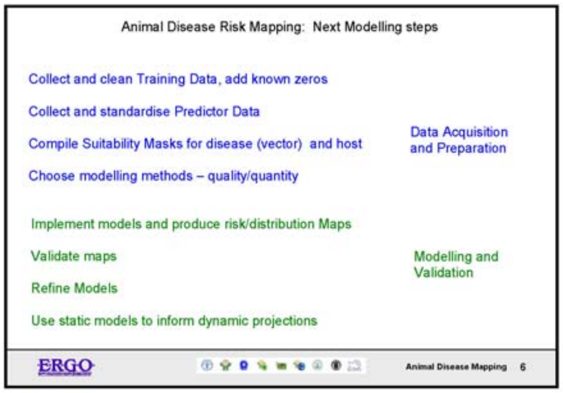

Animal Disease Risk Mapping: Basic Requirements
Inputs:
Disease Distributions
Animal Distributions
Modelling Methods
Predictor Archive
Possible Outputs:
Known Disease Maps
Potential Disease Risk Maps
Disease Spread Projections
ERGO
Animal Disease Mapping 1

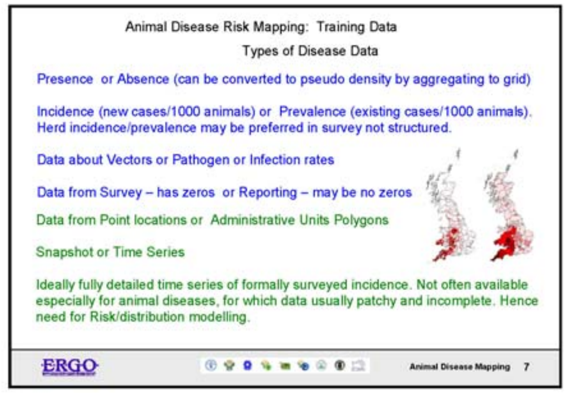

Animal Disease Risk Mapping: Basic Requirements
DRAFT
Input Presence
Output probability
Can also do abundance, and fill (internal) gaps ……..
ERGO
Animal Disease Mapping 2

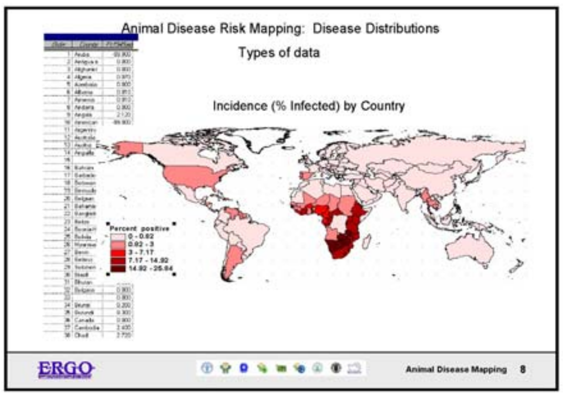

Animal Disease Risk Mapping: Basic Requirements
Observed
Predicted
ERGO
Animal Disease Mapping 3

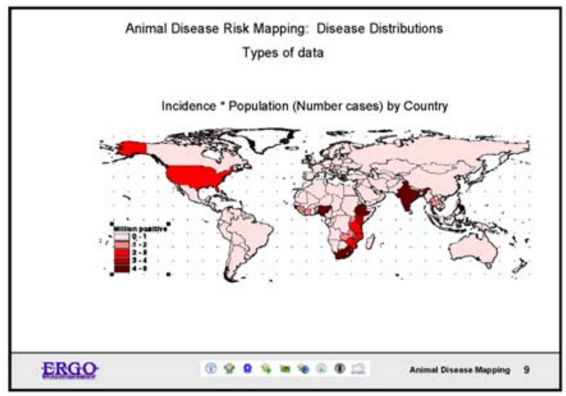

Animal Disease Risk Mapping: Modelling methods
Similar for Disease and Animals
Interpolation/Extrapolation (filling holes or expanding into unknown areas)
Weighting using rates and known host or population distributions
Cluster mapping (useful for control strategy development)
Process Based Modelling (need a lot of detailed input)
Statistical Modelling
Relate disease/animal density to predictors for which required resolution maps are available, and calculate distribution
Multiple Regression
Discriminant Analysis
Logistic Regression
Regression Trees
Maximum Entropy Models
Most require training data and predictor archive
Appropriate analysis resolution: depends on autocorrelation distance
ERGO
Animal Disease Mapping 4

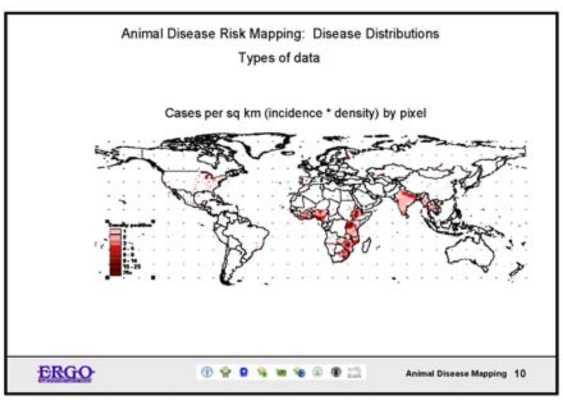

Animal Disease Risk Mapping: Modelling methods
Extract raster values for sample points
Basic technique: Establish statistical links between known presence/absence and a series of potential predictors for a large number of regularly spaced sample points (via logistic regression)
Known
Predictor 1
Calculate a 'regression equation' of the form: Known=A*predictor 1 + B*Predictor 2…+ constant)
NB there can be several predictors in the equations
First, convert all data maps to images with the same pixel size (resolution). Then extract values for each data type at fixed sample points (hatched squares). NB one of these maps must be of known values to be modelled.
Providing the equation is statistically significant (i.e. reliable), apply the right hand side of the equation to all the pixels in your images, not just the ones you sampled
Basis for multivariate analysis and distribution modelling
ERGO
Animal Disease Mapping 5

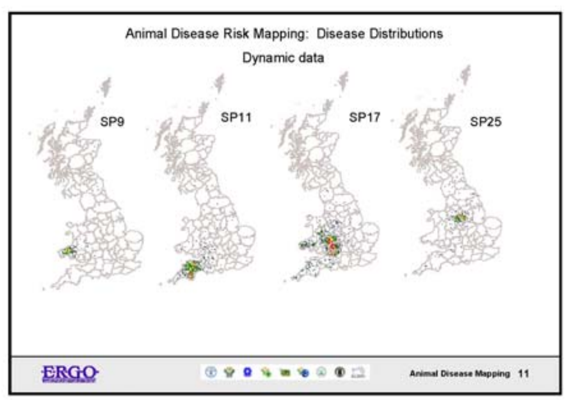

Animal Disease Risk Mapping: Next Modelling steps
Collect and clean Training Data, add known zeros
Collect and standardise Predictor Data
Compile Suitability Masks for disease (vector) and host
Choose modelling methods – quality/quantity
Data Acquisition and Preparation
Implement models and produce risk/distribution Maps
Validate maps
Refine Models
Use static models to inform dynamic projections
Modelling and Validation
ERGO
Animal Disease Mapping 6

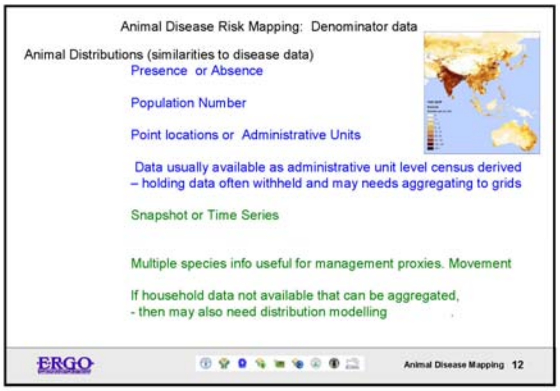

Animal Disease Risk Mapping: Training Data
Types of Disease Data
Presence or Absence (can be converted to pseudo density by aggregating to grid)
Incidence (new cases/1000 animals) or Prevalence (existing cases/1000 animals). Herd incidence/prevalence may be preferred in survey not structured.
Data about Vectors or Pathogen or Infection rates
Data from Survey – has zeros or Reporting – may be no zeros
Data from Point locations or Administrative Units Polygons
Snapshot or Time Series
Ideally fully detailed time series of formally surveyed incidence. Not often available especially for animal diseases, for which data usually patchy and incomplete. Hence need for Risk/distribution modelling.
ERGO
Animal Disease Mapping 7

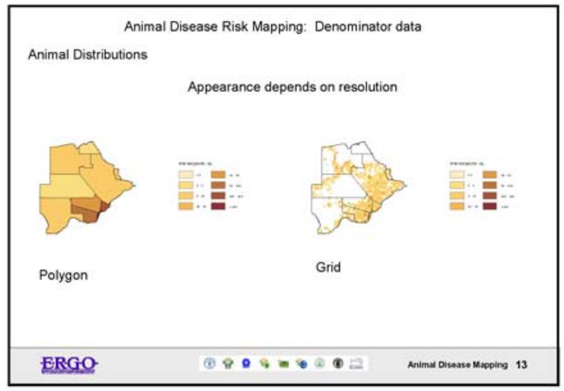

Animal Disease Risk Mapping: Disease Distributions
Types of data
Incidence (% Infected) by Country
Percent positive
0 - 0.82
0.82 - 3
3 - 7.17
7.17 - 14.92
14.92 - 25.04
ERGO
Animal Disease Mapping 8

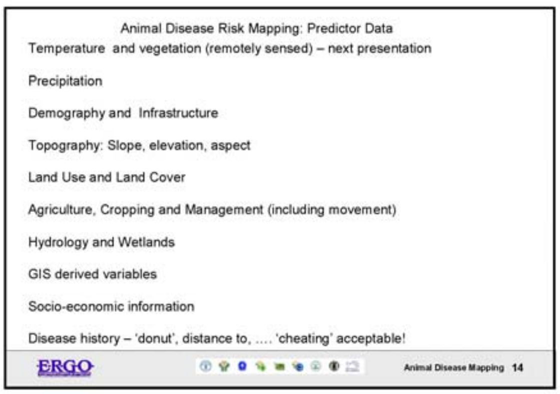

Animal Disease Risk Mapping: Disease Distributions
Types of data
Incidence * Population (Number cases) by Country
ERGO
Animal Disease Mapping 9

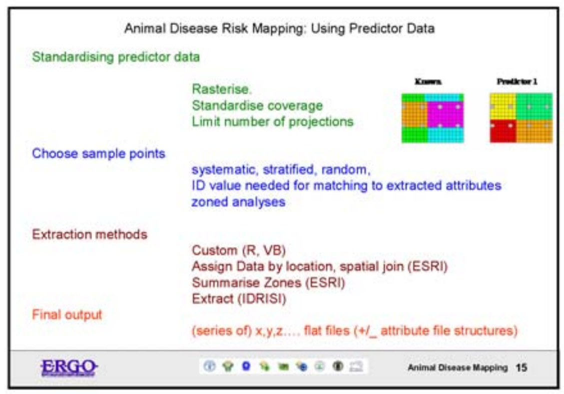

Animal Disease Risk Mapping: Disease Distributions
Types of data
Cases per sq km (incidence * density) by pixel
ERGO
Animal Disease Mapping 10

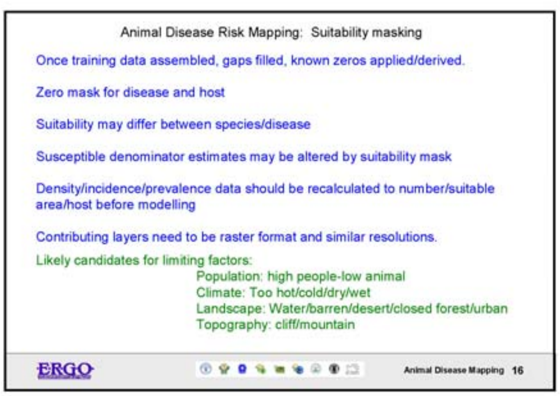

Animal Disease Risk Mapping: Disease Distributions
Dynamic data
SP9
SP11
SP17
SP25
ERGO
Animal Disease Mapping 11

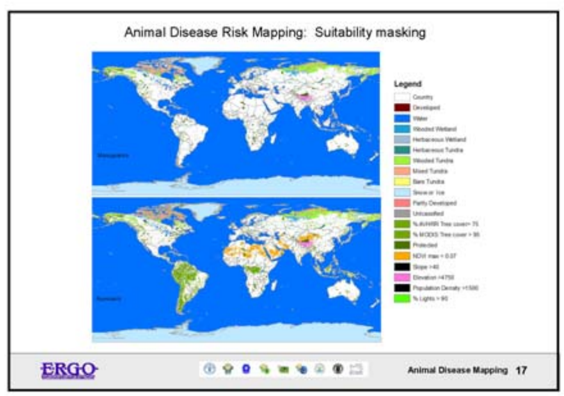

Animal Disease Risk Mapping: Denominator data
Animal Distributions (similarities to disease data)
Presence or Absence
Population Number
Point locations or Administrative Units
Data usually available as administrative unit level census derived – holding data often withheld and may needs aggregating to grids
Snapshot or Time Series
Multiple species info useful for management proxies. Movement
If household data not available that can be aggregated,
- then may also need distribution modelling
ERGO
Animal Disease Mapping 12

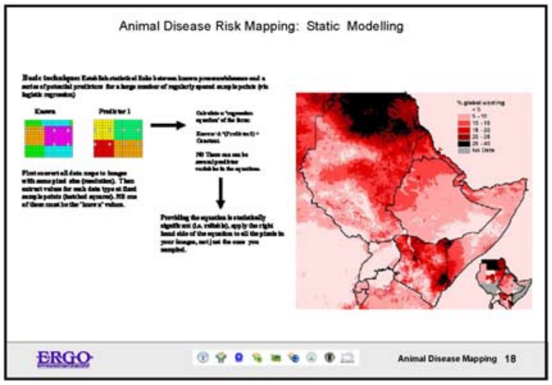

Animal Disease Risk Mapping: Denominator data
Animal Distributions
Appearance depends on resolution
Polygon
Grid
ERGO
Animal Disease Mapping 13

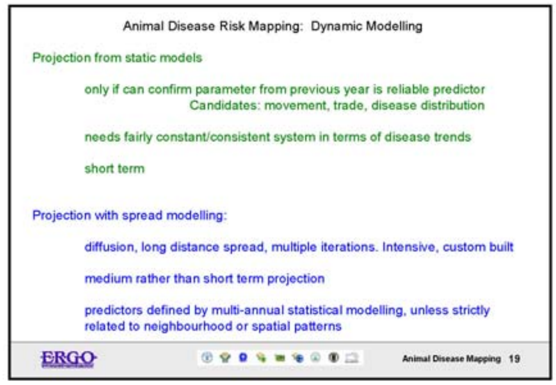

Animal Disease Risk Mapping: Predictor Data
Temperature and vegetation (remotely sensed) – next presentation
Precipitation
Demography and Infrastructure
Topography: Slope, elevation, aspect
Land Use and Land Cover
Agriculture, Cropping and Management (including movement)
Hydrology and Wetlands
GIS derived variables
Socio-economic information
Disease history – 'donut', distance to, ... 'cheating' acceptable!
ERGO
Animal Disease Mapping 14

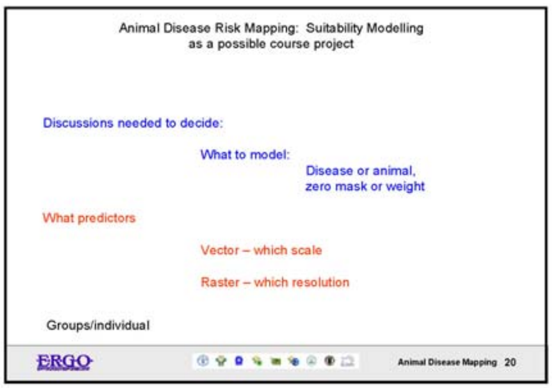

Animal Disease Risk Mapping: Using Predictor Data
Standardising predictor data
Rasterise.
Standardise coverage
Limit number of projections
Known
Predictor 1
Choose sample points
systematic, stratified, random,
ID value needed for matching to extracted attributes
zoned analyses
Extraction methods
Custom (R, VB)
Assign Data by location, spatial join (ESRI)
Summarise Zones (ESRI)
Extract (IDRISI)
Final output
(series of) x,y,z.... flat files (+/_ attribute file structures)
ERGO
Animal Disease Mapping 15

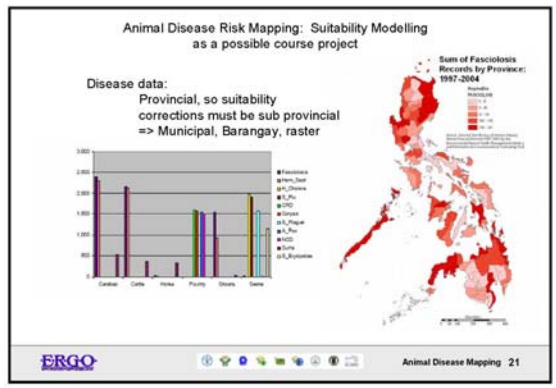

Animal Disease Risk Mapping: Suitability masking
Once training data assembled, gaps filled, known zeros applied/derived.
Zero mask for disease and host
Suitability may differ between species/disease
Susceptible denominator estimates may be altered by suitability mask
Density/incidence/prevalence data should be recalculated to number/suitable area/host before modelling
Contributing layers need to be raster format and similar resolutions.
Likely candidates for limiting factors:
Population: high people-low animal
Climate: Too hot/cold/dry/wet
Landscape: Water/barren/desert/closed forest/urban
Topography: cliff/mountain
ERGO
Animal Disease Mapping 16

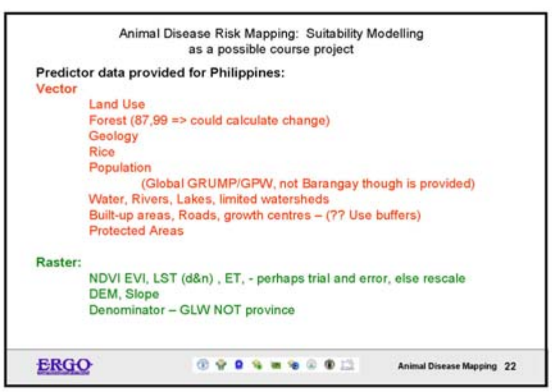

Animal Disease Risk Mapping: Suitability masking
Legend
Country
Developed
Water
Wooded Wetland
Herbaceous Wetland
Herbaceous Tundra
Wooded Tundra
Mixed Tundra
Bare Tundra
Snow or Ice
Partly Developed
Unclassified
Protected
ERGO
Animal Disease Mapping 17

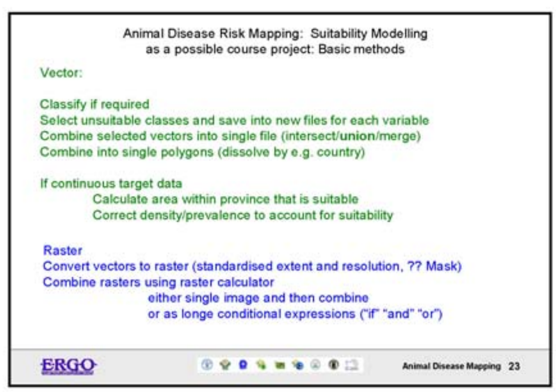

Animal Disease Risk Mapping: Static Modelling
ERGO
Animal Disease Mapping 18

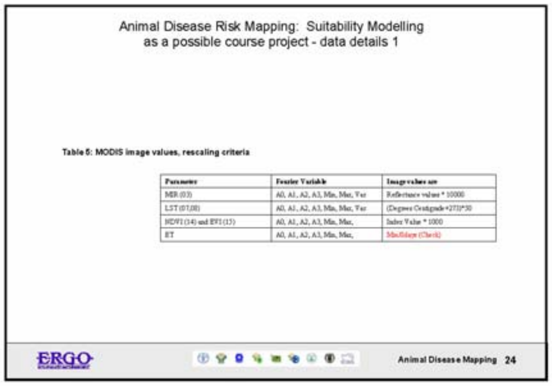

Animal Disease Risk Mapping: Dynamic Modelling
Projection from static models
only if can confirm parameter from previous year is reliable predictor
Candidates: movement, trade, disease distribution
needs fairly constant/consistent system in terms of disease trends
short term
Projection with spread modelling:
diffusion, long distance spread, multiple iterations. Intensive, custom built
medium rather than short term projection
predictors defined by multi-annual statistical modelling, unless strictly related to neighbourhood or spatial patterns
ERGO
Animal Disease Mapping 19


Animal Disease Risk Mapping: Suitability Modelling
as a possible course project
Discussions needed to decide:
What to model:
Disease or animal,
zero mask or weight
What predictors
Vector – which scale
Raster – which resolution
Groups/individual
ERGO
Animal Disease Mapping 20

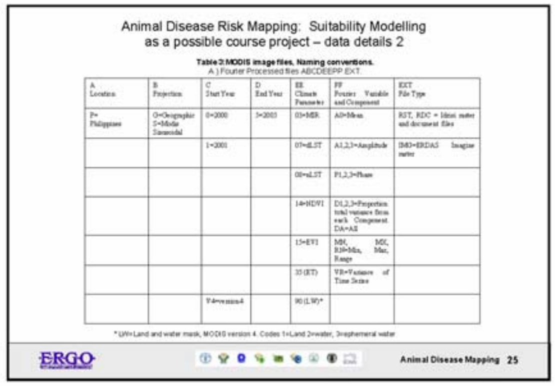

Animal Disease Risk Mapping: Suitability Modelling
as a possible course project
Disease data:
Provincial, so suitability
corrections must be sub provincial
=> Municipal, Barangay, raster
Sum of Fasciolosis Records by Province: 1997-2004
ERGO
Animal Disease Mapping 21


Animal Disease Risk Mapping: Suitability Modelling
as a possible course project
Predictor data provided for Philippines:
Vector
Land Use
Forest (87,99 => could calculate change)
Geology
Rice
Population
(Global GRUMP/GPW, not Barangay though is provided)
Water, Rivers, Lakes, limited watersheds
Built-up areas, Roads, growth centres – (?? Use buffers)
Protected Areas
Raster:
NDVI EVI, LST (d&n) , ET, - perhaps trial and error, else rescale
DEM, Slope
Denominator – GLW NOT province
ERGO
Animal Disease Mapping 22

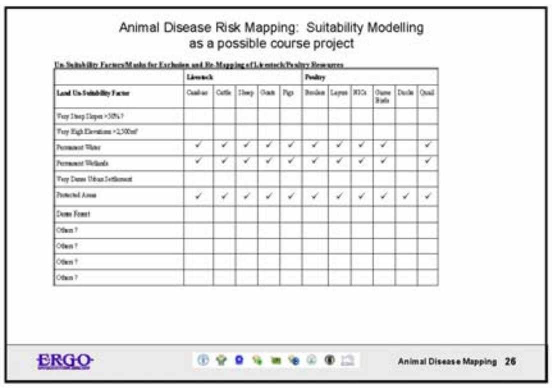

Animal Disease Risk Mapping: Suitability Modelling
as a possible course project: Basic methods
Vector:
Classify if required
Select unsuitable classes and save into new files for each variable
Combine selected vectors into single file (intersect/union/merge)
Combine into single polygons (dissolve by e.g. country)
If continuous target data
Calculate area within province that is suitable
Correct density/prevalence to account for suitability
Raster
Convert vectors to raster (standardised extent and resolution, ?? Mask)
Combine rasters using raster calculator
either single image and then combine
or as longe conditional expressions ("if" "and" "or")
ERGO
Animal Disease Mapping 23



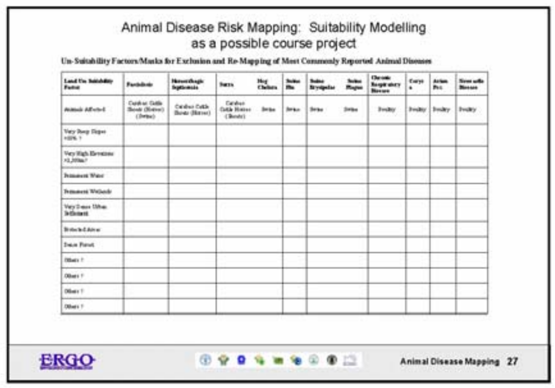
Parameter	Fourier Variable	Image values are
MIR (03)	A0, A1, A2, A3, Min, Max, Var	Reflectance values * 10000
LST (07,08)	A0, A1, A2, A3, Min, Max, Var	(Degrees Centigrade +273)*50
NDVI (14) and EVI (15)	A0, A1, A2, A3, Min, Max,	Index Value * 1000
ET	A0, A1, A2, A3, Min, Max,	Mm/day (Check)



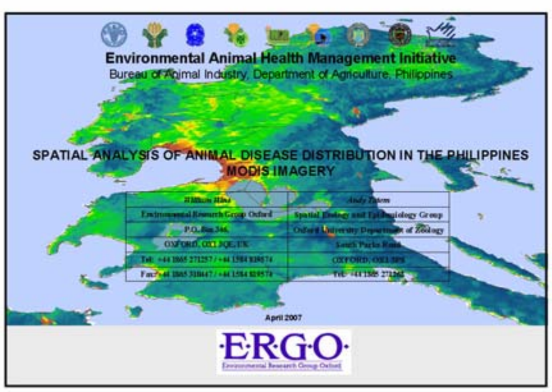
A Location	B Projection	C Start Year	D End Year	EE Climate Parameter	PP Fourier Variable and Component	EXT File Type
P= Philippines	G=Geographic S=Modis Sinusoidal	0=2000	3=2003	03=MIR	A0=Mean	RST, RDC = Idrisi raster and document files
		1=2001		07=dLST	A1,2,3=Amplitude	IMG=ERDAS Imagine raster
				08=nLST	P1,2,3=Phase	
				14=NDVI	D1,2,3=Proportion total variance from each Component DA=All	
				15=EVI	MN, MX, RN=Min, Max, Range	
				35 (ET)	VR=Variance of Time Series	
		V4=version4		90 (LW)*		

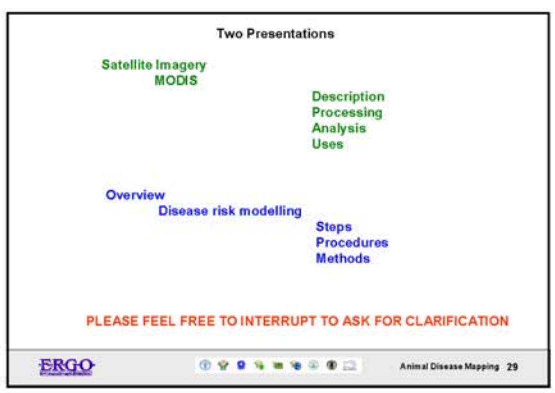


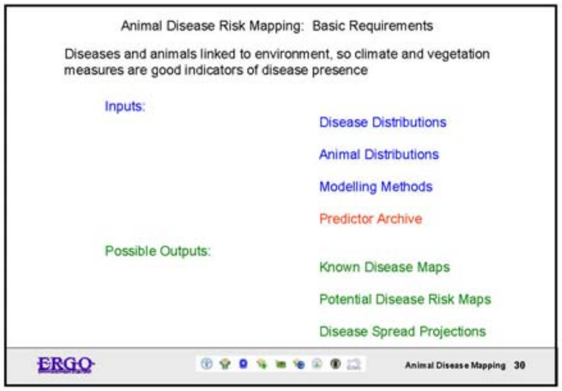
	Livestock					Poultry					
Land Un-Suitability Factor	Carabao	Cattle	Sheep	Goats	Pigs	Broilers	Layers	NICs	Game Birds	Ducks	Quail
Very Steep Slopes >50%?											
Very High Elevations >2,500m?											
Permanent Water	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Permanent Wetlands	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Very Dense Urban Settlement											
Protected Areas	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Dense Forest											
Others ?											
Others ?											
Others ?											
Others ?											

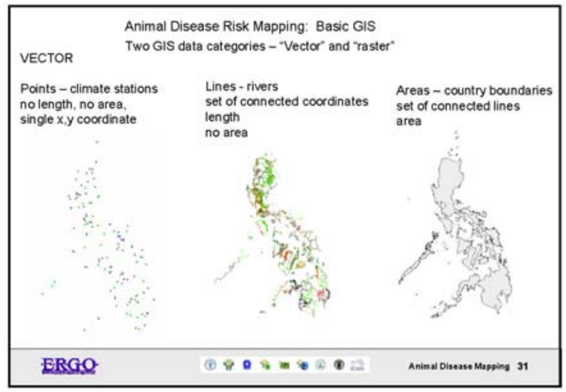


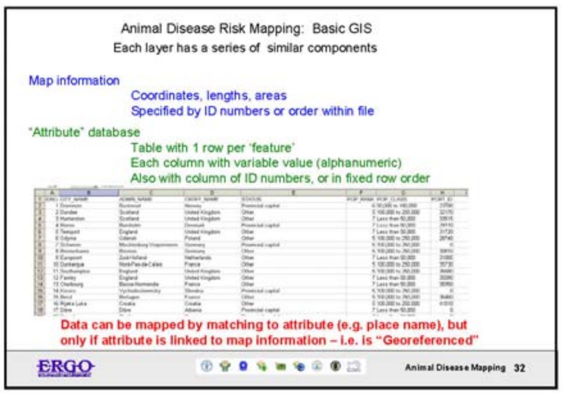
Land Use Suitability Factor	Fasciolosis	Haemorrhagic Septicemia	Surra	Hog Cholera	Swine Flu	Swine Erysipelas	Swine Plague	Chronic Respiratory Disease	Coryza	Avian Pox	Newcastle Disease
Animals Affected	Carabao Cattle Goats (Horses) (Swine)	Carabao Cattle Goats (Horses)	Carabao Cattle Horses (Goats)	Swine	Swine	Swine	Swine	Poultry	Poultry	Poultry	Poultry
Very Steep Slopes >50% ?											
Very High Elevations >2,500m?											
Permanent Water											
Permanent Wetlands											
Very Dense Urban Settlement											
Protected Areas											
Dense Forest											
Others ?											
Others ?											
Others ?											
Others ?											

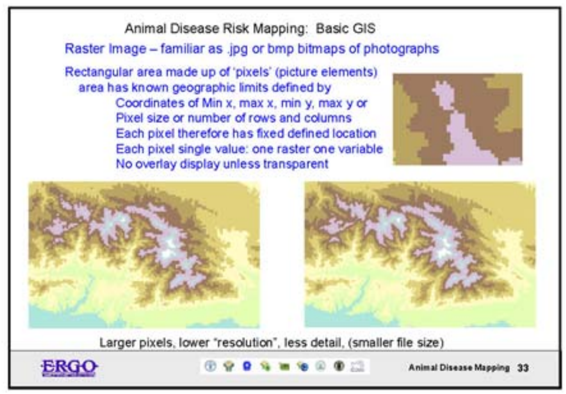


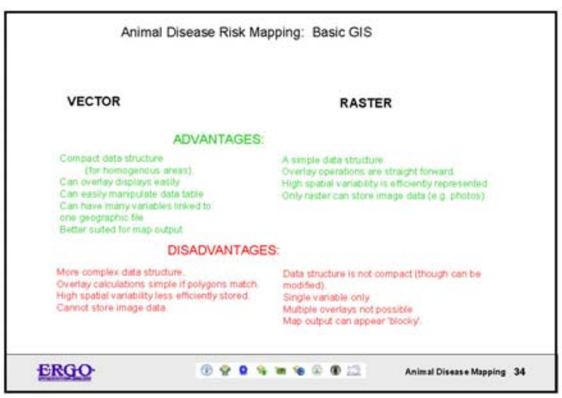


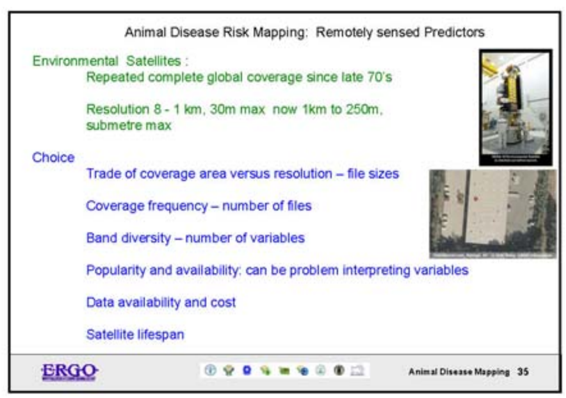

Animal Disease Risk Mapping: Basic Requirements
Diseases and animals linked to environment, so climate and vegetation measures are good indicators of disease presence
Inputs:
Disease Distributions
Animal Distributions
Modelling Methods
Predictor Archive
Possible Outputs:
Known Disease Maps
Potential Disease Risk Maps
Disease Spread Projections
ERGO
Animal Disease Mapping 30

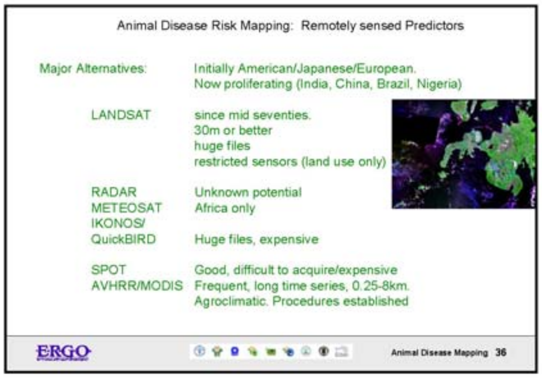

Animal Disease Risk Mapping: Basic GIS
Two GIS data categories – "Vector" and "raster"
VECTOR
Points – climate stations
no length, no area,
single x,y coordinate
Lines - rivers
set of connected coordinates
length
no area
Areas – country boundaries
set of connected lines
area
ERGO
Animal Disease Mapping 31

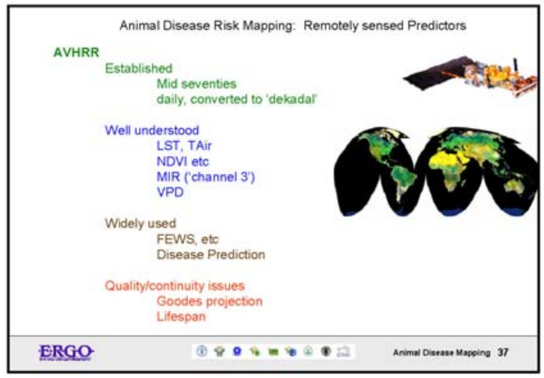

Animal Disease Risk Mapping: Basic GIS
Each layer has a series of similar components
Map information
Coordinates, lengths, areas
Specified by ID numbers or order within file
"Attribute" database
Table with 1 row per 'feature'
Each column with variable value (alphanumeric)
Also with column of ID numbers, or in fixed row order
Data can be mapped by matching to attribute (e.g. place name), but only if attribute is linked to map information – i.e. is "Georeferenced"
ERGO
Animal Disease Mapping 32

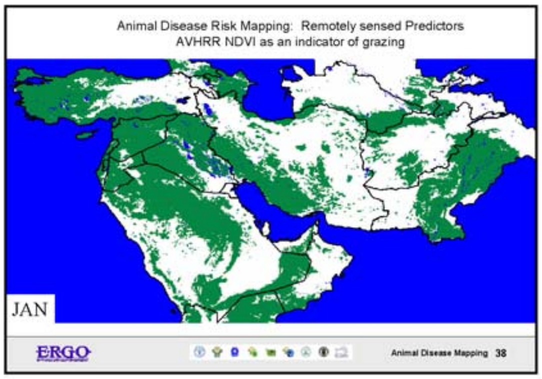

Animal Disease Risk Mapping: Basic GIS
Raster Image – familiar as .jpg or bmp bitmaps of photographs
Rectangular area made up of 'pixels' (picture elements)
area has known geographic limits defined by
Coordinates of Min x, max x, min y, max y or
Pixel size or number of rows and columns
Each pixel therefore has fixed defined location
Each pixel single value: one raster one variable
No overlay display unless transparent
Larger pixels, lower "resolution", less detail, (smaller file size)
ERGO
Animal Disease Mapping 33

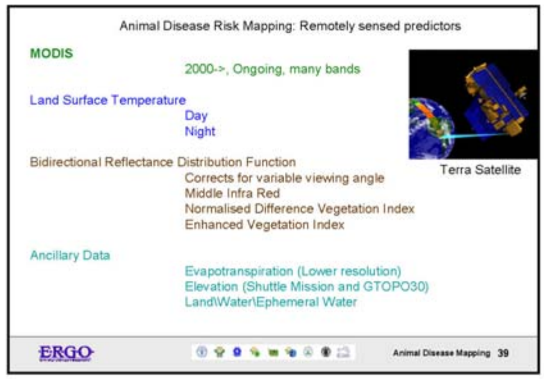

Animal Disease Risk Mapping: Basic GIS
VECTOR
RASTER
ADVANTAGES:
Compact data structure
(for homogenous areas)
Can overlay displays easily
Can easily manipulate data table
Can have many variables linked to one geographic file
Better suited for map output
A simple data structure
Overlay operations are straight forward
High spatial variability is efficiently represented
Only raster can store image data (e.g. photos)
DISADVANTAGES:
More complex data structure.
Overlay calculations simple if polygons match
High spatial variability less efficiently stored
Cannot store image data
Data structure is not compact (though can be modified).
Single variable only
Multiple overlays not possible
Map output can appear 'blocky'.
ERGO
Animal Disease Mapping 34

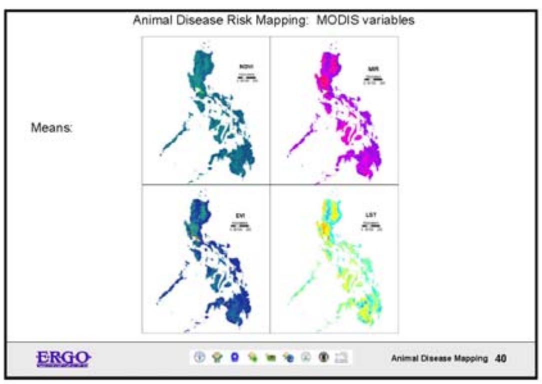

Animal Disease Risk Mapping: Remotely sensed Predictors
Environmental Satellites :
Repeated complete global coverage since late 70's
Resolution 8 - 1 km, 30m max now 1km to 250m, submetre max
Choice
Trade of coverage area versus resolution – file sizes
Coverage frequency – number of files
Band diversity – number of variables
Popularity and availability: can be problem interpreting variables
Data availability and cost
Satellite lifespan
ERGO
Animal Disease Mapping 35

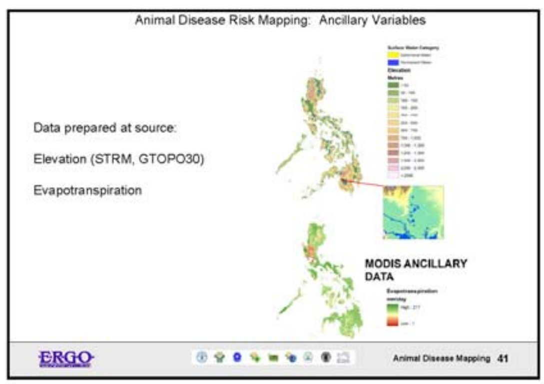

Animal Disease Risk Mapping: Remotely sensed Predictors
Major Alternatives:
Initially American/Japanese/European.
Now proliferating (India, China, Brazil, Nigeria)
LANDSAT
since mid seventies.
30m or better
huge files
restricted sensors (land use only)
RADAR
METEOSAT
IKONOS/
QuickBIRD
Unknown potential
Africa only
Huge files, expensive
SPOT
AVHRR/MODIS
Good, difficult to acquire/expensive
Frequent, long time series, 0.25-8km.
Agroclimatic. Procedures established
ERGO
Animal Disease Mapping 36

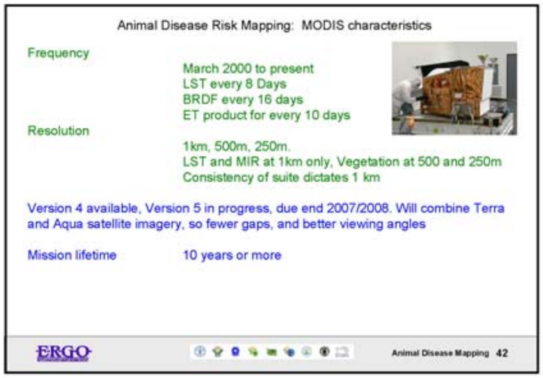

Animal Disease Risk Mapping: Remotely sensed Predictors
AVHRR
Established
Mid seventies
daily, converted to 'dekadal'
Well understood
LST, TAir
NDVI etc
MIR ('channel 3')
VPD
Widely used
FEWS, etc
Disease Prediction
Quality/continuity issues
Goodes projection
Lifespan
ERGO
Animal Disease Mapping 37

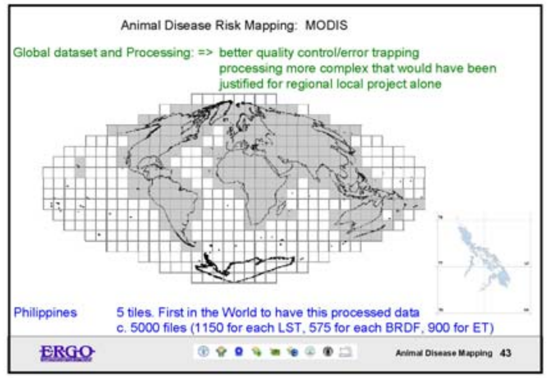

Animal Disease Risk Mapping: Remotely sensed Predictors
AVHRR NDVI as an indicator of grazing
JAN
ERGO
Animal Disease Mapping 38

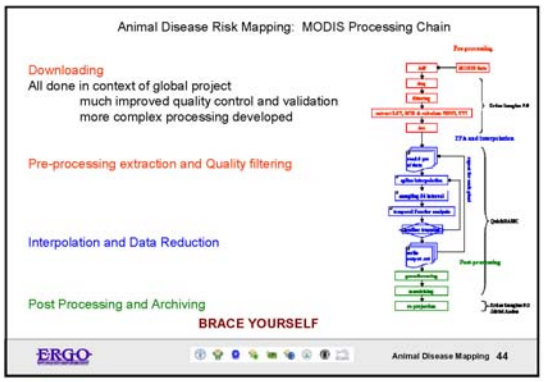

Animal Disease Risk Mapping: Remotely sensed predictors
MODIS
2000->, Ongoing, many bands
Land Surface Temperature
Day
Night
Bidirectional Reflectance Distribution Function
Corrects for variable viewing angle
Middle Infra Red
Normalised Difference Vegetation Index
Enhanced Vegetation Index
Terra Satellite
Ancillary Data
Evapotranspiration (Lower resolution)
Elevation (Shuttle Mission and GTOPO30)
Land\Water\Ephemeral Water
ERGO
Animal Disease Mapping 39

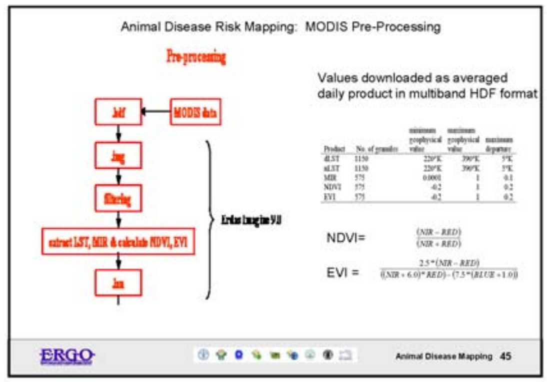

Animal Disease Risk Mapping: MODIS variables
Means:
NDVI
MIR
EVI
LST
ERGO
Animal Disease Mapping 40

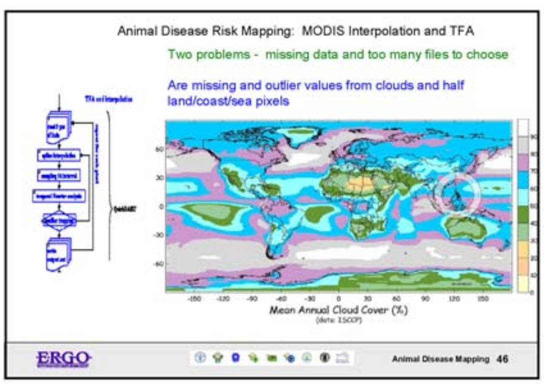

Animal Disease Risk Mapping: Ancillary Variables
Data prepared at source:
Elevation (STRM, GTOPO30)
Evapotranspiration
MODIS ANCILLARY DATA
ERGO
Animal Disease Mapping 41

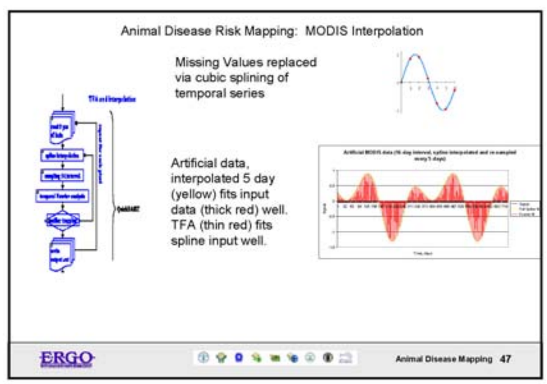

Animal Disease Risk Mapping: MODIS characteristics
Frequency
March 2000 to present
LST every 8 Days
BRDF every 16 days
ET product for every 10 days
Resolution
1km, 500m, 250m.
LST and MIR at 1km only, Vegetation at 500 and 250m
Consistency of suite dictates 1 km
Version 4 available, Version 5 in progress, due end 2007/2008. Will combine Terra and Aqua satellite imagery, so fewer gaps, and better viewing angles
Mission lifetime
10 years or more
ERGO
Animal Disease Mapping 42

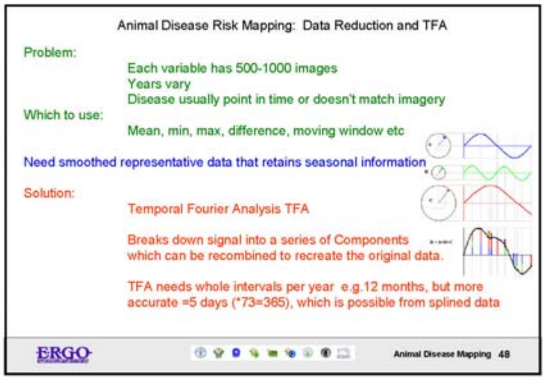

Animal Disease Risk Mapping: MODIS
Global dataset and Processing: => better quality control/error trapping
processing more complex that would have been justified for regional local project alone
Philippines
5 tiles. First in the World to have this processed data
c. 5000 files (1150 for each LST, 575 for each BRDF, 900 for ET)
ERGO
Animal Disease Mapping 43

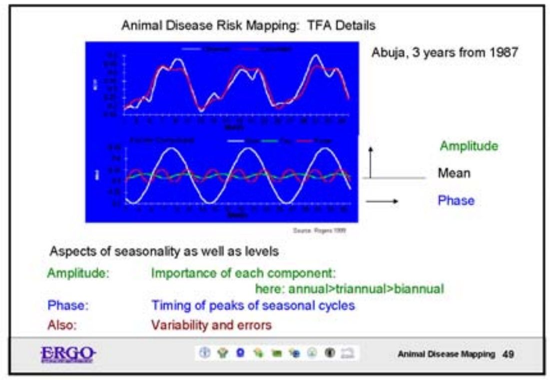

Animal Disease Risk Mapping: MODIS Processing Chain
Downloading
All done in context of global project
much improved quality control and validation
more complex processing developed
Pre-processing extraction and Quality filtering
Interpolation and Data Reduction
Post Processing and Archiving
BRACE YOURSELF
ERGO
Animal Disease Mapping 44

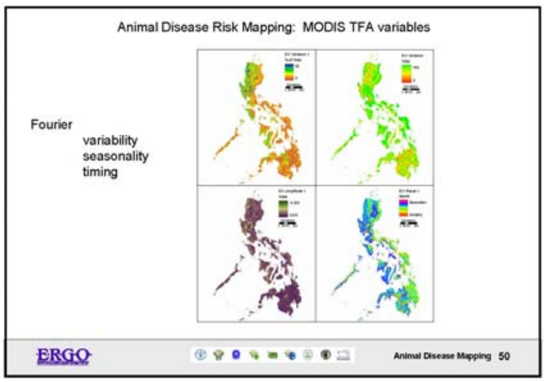

Animal Disease Risk Mapping: MODIS Pre-Processing
Pre-processing
MODIS data
extract LST, MIR & calculate NDVI, EVI
Values downloaded as averaged daily product in multiband HDF format
NDVI=
EVI =
ERGO
Animal Disease Mapping 45

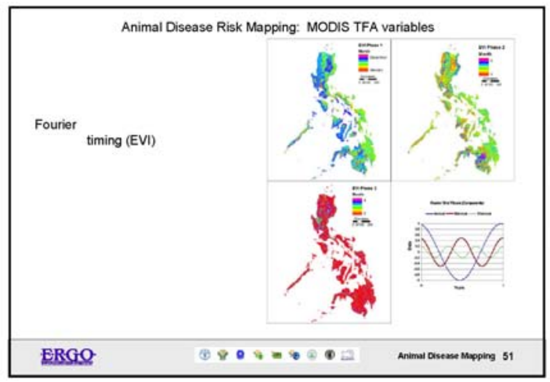

Animal Disease Risk Mapping: MODIS Interpolation and TFA
Two problems - missing data and too many files to choose
Are missing and outlier values from clouds and half land/coast/sea pixels
Mean Annual Cloud Cover (%)
ERGO
Animal Disease Mapping 46

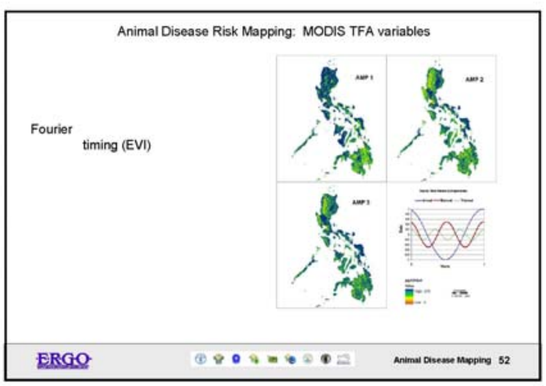

Animal Disease Risk Mapping: MODIS Interpolation
Missing Values replaced via cubic splining of temporal series
Artificial data, interpolated 5 day (yellow) fits input data (thick red) well. TFA (thin red) fits spline input well.
ERGO
Animal Disease Mapping 47

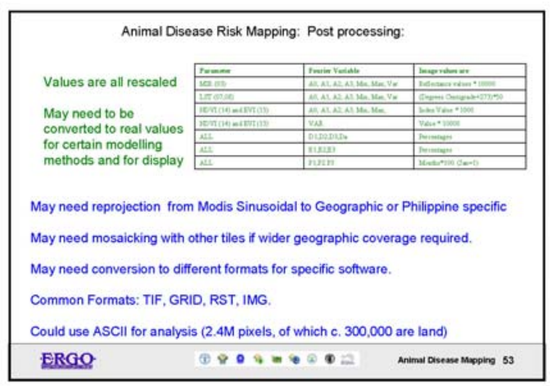

Animal Disease Risk Mapping: Data Reduction and TFA
Problem:
Each variable has 500-1000 images
Years vary
Disease usually point in time or doesn't match imagery
Which to use:
Mean, min, max, difference, moving window etc
Need smoothed representative data that retains seasonal information
Solution:
Temporal Fourier Analysis TFA
Breaks down signal into a series of Components which can be recombined to recreate the original data.
TFA needs whole intervals per year e.g.12 months, but more accurate =5 days (*73=365), which is possible from splined data
ERGO
Animal Disease Mapping 48

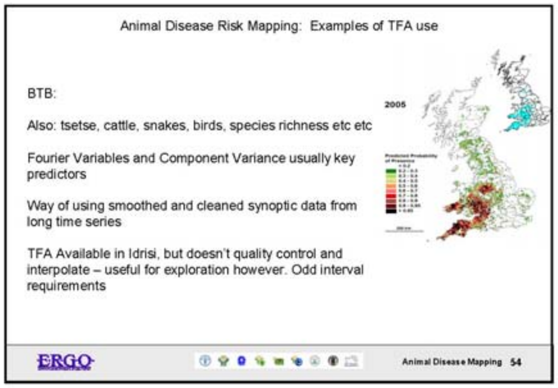

Animal Disease Risk Mapping: TFA Details
Abuja, 3 years from 1987
Amplitude
Mean
Phase
Aspects of seasonality as well as levels
Amplitude: Importance of each component:
here: annual>triannual>biannual
Phase: Timing of peaks of seasonal cycles
Also: Variability and errors
ERGO
Animal Disease Mapping 49

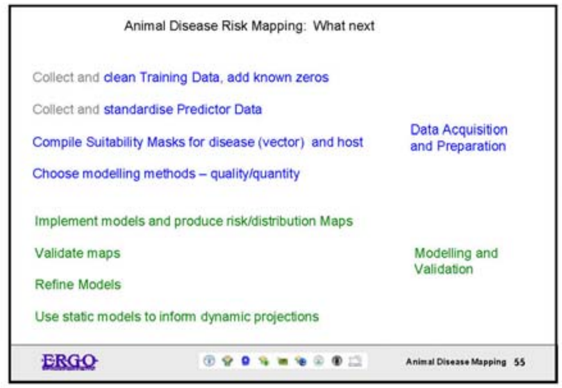

Animal Disease Risk Mapping: MODIS TFA variables
Fourier
variability
seasonality
timing
ERGO
Animal Disease Mapping 50


Animal Disease Risk Mapping: MODIS TFA variables
Fourier
timing (EVI)
ERGO
Animal Disease Mapping 51


Animal Disease Risk Mapping: MODIS TFA variables
Fourier
timing (EVI)
AMP 1
AMP 2
AMP 3
ERGO
Animal Disease Mapping 52


Animal Disease Risk Mapping: Post processing:
Values are all rescaled
May need to be converted to real values for certain modelling methods and for display
Parameter
Fourier Variable
Image values are
May need reprojection from Modis Sinusoidal to Geographic or Philippine specific
May need mosaicking with other tiles if wider geographic coverage required.
May need conversion to different formats for specific software.
Common Formats: TIF, GRID, RST, IMG.
Could use ASCII for analysis (2.4M pixels, of which c. 300,000 are land)
ERGO
Animal Disease Mapping 53


Animal Disease Risk Mapping: Examples of TFA use
BTB:
Also: tsetse, cattle, snakes, birds, species richness etc etc
Fourier Variables and Component Variance usually key predictors
Way of using smoothed and cleaned synoptic data from long time series
TFA Available in Idrisi, but doesn't quality control and interpolate – useful for exploration however. Odd interval requirements
2005
ERGO
Animal Disease Mapping 54


Animal Disease Risk Mapping: What next
Collect and clean Training Data, add known zeros
Collect and standardise Predictor Data
Compile Suitability Masks for disease (vector) and host
Choose modelling methods – quality/quantity
Data Acquisition and Preparation
Implement models and produce risk/distribution Maps
Validate maps
Refine Models
Use static models to inform dynamic projections
Modelling and Validation
ERGO
Animal Disease Mapping 55