Biodiversity Data Discovery through Data Mining in Mountain areas

Priyanka Verma

Department of Computer Science and IT, The IIS University, Jaipur

Abstract: The degree of variation of life forms within a given species, ecosystem, biome, or an entire planet is biodiversity. Mining biodiversity databases of mountain organisms will help us to get better understanding of mountain biodiversity. The data will be taken from georeference biodiversity databases created through global biodiversity information facility. Biodiversity for particular mountain area depends upon range of elevation, associated climatic trends, topographic and soil peculiarities, fragmentation and connectivity amongst biota. Data discovered through mining databases will be categorized according to major taxonomic group with more coverage on animals specially birds, mammals and fishes.Biodiverstity data gathered will help us in solving applied issues such as finding endangered species, migration of species, invasion of new species, conservation planning etc., societal issues such as ecotourism, recreation, public health and also basic issues which includes taxonomy, diversity, ecology and evolution. Using biodiversity databases helps us in managing biodiversity, test ecological and evolutionary theories measure impact of climate change on various species and its effects on conservation efforts. Data will be analyzed through niche modeling which will help us to explain past and future trends in mountain biodiversity. Niche of species is defined as set of ecological conditions within which it is able to maintain population without immigration. In this paper we will discover niche of species through mining databases.

Keywords: Georefernce, Topographic, Fragmentation, Taxonomy, etc.

I. INTRODUCTION

Existing and emerging electronic databases are amongst the most promising tools for accessing mountain biodiversity. Mountain biodiversity is primarily affected by gradients of altitude and associated climatic trends, topographic and soil peculiarities, and fragmentation and connectivity amongst biota. The mountains of the world exhibit different climatic trends along their slopes, with only few factors, such as the decline in atmospheric pressure, ambient temperature, and clear sky radiation changing in a common, altitude-specific way across the globe. None of the other key components of climate, such as cloudiness and, with it, actual solar radiation or precipitation and associated soil moisture, show such global trends, and hence are not altitude-specific. The separation of global from regional environmental conditions along elevational transects offers new perspectives for understanding adaptation of mountain biota. For studying mountain biodiversity we concentrated on The Hindu Kush Himalayan region which is significantly rich in terms of biodiversity resources but it presents us with one of the least studied in the globe. The available data in the region are inaccessible and not well managed and formatted. So we can say that the inventory, assessment and sharing well documented biodiversity information of the region have become essential to improve understanding, efficient conservation and management of these resources. "For publishing, harvesting and using biodiversity data from the HKH region Global Mountain Biodiversity Assessment

(GMBA), which aims to document and synthesize knowledge on the biological richness of the mountains of the world and the changes undergoing as a result of direct and indirect human along with Global Biodiversity Information Facility (GBIF) organized a workshop"(Workshop convened by ICIMOD,2008). The workshop brought together 25 representatives from eight regional member countries of the Hindu Kush Himalayan region.It focused on enhancing awareness of the central role of georeferencing in biodiversity database building. Once achieved, this permits linkage of biological information with other geophysical information, particularly climate data which changes with altitude change. Objectives of the workshop were biodiversity data discovery, publishing and use for the HKH countries and the region; and to strengthen the capacity of biodiversity researchers and data publishers from the HKH region to discover, digitize and publish biodiversity data by adopting GBIF promoted tools,

standards and processes. "It aims to document and synthesize knowledge on the biological richness of the mountains of the world and the changes undergoing as a result of direct and indirect human influences." (Workshop convened by ICIMOD,2008)

II. OPEN ACCESS TO BIODIVERSITY DATA THROUGH GBIF

GBIF is an intergovernmental initiative to share biodiversity information across borders. The institution currently has 54 countries and 44 intergovernmental organizations as members globally. It was established in 2000. It is government initiated and funded in response to the needs of government agency for biodiversity information access and management. It facilitates free and open access to biodiversity data worldwide via internet, for conservation and sustainable development. GBIF has the following objectives and principles:

- Openness
- Online data sharing of more than 200 million biodiversity data records
- Facilitating access/exchange of high quality data.

"GBIF has a mission to make the world's primary data on biodiversity freely and universally available via the Internet"(Workshop convened by ICIMOD,2008).GBIF is a preferred source mechanism for biodiversity data and provides a platform for publishing, integration, access, and use of biodiversity data.GBIF has already established networks, data exchange standards, and an information architecture that enables interoperability and facilitates mining of biodiversity data.

A. Format followed for providing open access to data

XML forms the basis for information sharing. It is a freeform text document with content which is an ideal format for online information and data transfer. It is a semi-structured database which can be used for books, documents, worksheets or databases. Hypertext Markup Language (HTML) is an applied example of XML in a web browser. Keyhole Markup Language (KML), Geography Markup Language (GML) designed for maps and spatial information and Ecologically Markup Language (EML) for ecological data are other formats used

B. *Tools for publishing biodiversity data online*

There are lots of standards and different tools for publishing biodiversity data. Given below are the various wrapper tools for data publishing.

- Distributed Generic Information Retrieval (DiGIR) Wrapper Tool (2002) uses HTTP as a transport mechanism and XML for encoding message sent between the client and institution, allowing data in an online database to be exchanged in a standard format
- Biological Collection Access Services (BioCASE) Wrapper Tool (2003) builds on work started by DiGIR which also uses HTTP as a transport

mechanism and XML for encoding messages sent between client and the institution.
- Taxonomic Database Working Group (TDWG) Access Protocol for Information Retrieval (TAPIR) Wrapping Tool (2007) is designed as a generic tool to apply in domains other than biodiversity and natural science collections data. It combines and extends features of DiGIR and BioCASE protocols which are flexible to use the data exchange standards. ("Open Access to, and publication of, Mountain Biodiversity Data of the Hindu Kush Himalayan region ICIMOD", Sunita Chaudhary et al (2010))

C. Initiatives taken by GBIF to promote easy and open access to biodiversity data in HKH region

To develop a regional framework and partnership needed for promoting easy and open access to standardized and harmonized biodiversity information in the HKH region, a way forward has been sought by ICIMOD and GBIF The way forward is to:

- Activate ICIMOD as a regional node of GBIF: As a regional node, ICIMOD will provide technical support to the regional member countries as per required. The centre will also conduct national level trainings to the government officials, biodiversity researchers and managers by developing training materials and guidelines. A call for mini grants, proposals and awards as bridging fund will be made in order to help the national and regional partners to initiate or collaborate in biodiversity informatics project. Participation of national partners, scientists and researchers will be encouraged to support ICIMOD as a regional node of GBIF.
- Initiate a regional collaboration to develop and share biodiversity information in the HKH region by developing a concept proposal through regional consensus, developing collaborative proposal for funding with national partners and GMBA and implementing pilot and complementary project

III. DATA MINING OF BIODIVERSITY DATA

Openly accessible, interconnected electronic databases are used for scientific biodiversity research by performing data mining and data linking. Data mining is performed by combining data from phylogenetic and phylogeographic databases, regional species lists, classify them by elevation (e.g. selection of alpine species), geographic distribution and species range limits, also gather information on resilience of a species to change (life form, life cycle characteristics, reproduction, and phenological data).

A. Data Sources for data mining

Various mountain Biodiversity portals have been developed by GMBA in collaboration with GBIF It allows

exploring biodiversity archive data for mountain regions. Data can be searched from database by range of elevational or thermal belts. Various elevational belts are(PeterSaundry,2011)

- The montane belt extends from the lower mountain limit to the upper thermal limit of forest (irrespective of whether forest is currently present or not).
- The alpine belt is the temperature-driven treeless region between the natural climatic forest limit and the snowline that occurs worldwide.
- The nival belt is the terrain above the snowline, which is defined as the lowest elevation where snow is commonly present all year round (though not necessarily with full cover).
- The treeline ecotone is the transition zone between the montane and alpine belts. ("Creative Use of Mountain Biodiversity Databases ",Christian Körner et al(2002))

One such portal is "HKH Conservation Portal which provides free and universal access to biodiversity data. It is a thematic portal to promote sharing conservation information in the HKH. It aims to provide free and open access to primary data and information related to landscape-level conservation initiatives in the HKH region including PA, corridors and biodiversity resources"(http://www.icimod.org/hkhconservationportal).

B. Classification of biodiversity data

Biodiversity data in particular region can be classified as

Individual based data (primary occurrences, an individual at a place at a particular time)

Taxon based data (biological taxon characteristics, such as morphology, physiology, phylogeny, ecology, genetics) with more coverage on animals especially birds, mammals and fishes (i.e. Animalia; 72%, Protozoa: 1%, Plantae: 26% and Fungi: 1%).

Various attributes considered while classifying biodiversity are given below

- Plants: Biological attributes such as size (height), life form, flower features, current phenology, seed size, growth form, and other special attributes. These data can sometimes be obtained from taxonomic sources and stored in relational databases.
- Animals: Biological attributes such as size (width, length, etc), tropic habit, and interactions (prey, mutualistic species, host, phenology, life stage).
- Abundance or frequency measures (e.g. random sample of quadrants). Information on rareness, conservation status, dominant associates, population structure, if available.

A full, best-practice database entry should include the following types of data:

- Organism data (conventional taxonomic information)
- Geoinformation (coordinates, altitude)
- Habitat information (topographic, atmospheric)
- Date and time of observation, collection and recording
- Reference to a voucher or archive code
- Name of collector, observer, and recorder

Metadata that provide information on data sets, such as content, extent, accessibility, currency, completeness, accuracy, uncertainties, fitness for purpose and suitability for use, and enable the use of data by third parties without reference to the originator("Principles of Data Quality version 1.0",Chapman AD(2005))

C. Benefits of mining biodiversity data

The primary biodiversity data (PBD) discovered through data mining on online data is useful for answering following question such as

- How did that particular species arise?
- What is the contribution of mountain biodiversity to ecosystem integrity?
- What are the socioeconomic impacts on mountain biodiversity
- What is effect of environmental change on mountain biodiversity

These questions when answered will help in making predictions for solving applied issues (endangered species, migration, invasion, conservation planning, genetics etc), societal issues (ecotourism, recreation, public health) and basic issues (taxonomy, diversity, population dynamics, biogeography, ecology and evolution).Predictions are necessary as it fills data gaps and can provide reliable and transparent scenarios of the future.("An international framework to promote access to data", Arzberger P et al(2004))

i. Generation, evolution, Assembly of mountain biodiversity

The origin and assembly of mountain biota have to be understood by answering questions given below

Where did its taxa arise?

- How were taxa assembled over time?
- How many of the extant species resulted from the radiation of lineages that evolved within the area as opposed to the radiation of lineages that were introduced from other areas or even continents or other ecosystems?
- How important has long distance dispersal been for the assembly of mountain biota, and how and when did evolutionary lineages migrate from one mountain area to others?

- What are the main sources of long-distance dispersal events?
- Has the capacity of long-distance dispersal itself been a factor in the rapid radiation of alpine lineages?

Mountains are islands of varying size, and thus present a good opportunity to ask questions about genesis of mountain biota, the impact of competition from other biota on speciation rates, and adaptive evolution. Mountains have acted as refugee for species survival during extreme climatic events, including for ancient phylogenetic lineages.

ii. Contribution of mountain biodiversity to ecosystem integrity

Ecosystem integrity on Steep Mountain slopes and in high-elevation landscapes is mainly a question of soil stability, which in turn depends on plant cover. The insurance hypothesis of biodiversity suggests that the more diversity (e.g. genetic diversity, morph types) there is, the less likely it is that extreme events or natural diseases will lead to a decline in ecosystem functioning or a failure of vegetation to prevent soil erosion. "In steep terrain, more than anywhere else, catchments quality is intimately linked to ecosystem integrity. The provision of sustainable and clean supplies of water is the most important and increasingly limiting mountain resource" (PeterSaundry,2011).

To understand effect of mountain biodiversity to ecosystem integrity we need to answer following questions

- What is the contribution of mountain biodiversity to ecosystem integrity, i.e. slope stability?
- What is the functional redundancy in traits among organisms in a given area, what is their sensitivity to stress and disturbance (insect outbreaks, avalanches)?

These questions can be answered through data mining on Old vs. new inventory data, recent loss or gain of certain plant functional types (e.g. trees). Recent land cover change (remote sensing evidence, NDVI). Apart from information on composition of vegetation and functional traits of taxa (e.g. rooting depth, root architecture, growth form), geo-graphical information is needed (geomorphology: slope, relief, soil depth; climate, precipitation, evapotranspiration, extreme rain events, snow cover duration Comparison of different mountain regions (e.g. presence/absence of woody/non-woody vegetation). Spatial land cover information can be used to develop scenarios at landscape scale.

iii. Socioeconomic impacts on mountain biodiversity

Of all global change effects, land use is the predominant driver of changes in mountain biodiversity. By comparing areas of historically contrasting land use regimes we can learn how these human activities shape biota. Ratios of wilderness biodiversity to adjacent managed biodiversity indicate the actual impact of land use. The abundance of red list taxa or medicinal plants can be related to human population pressure and land use intensity. Humans shape mountain vegetation by clearing land, grazing, abandoning, collecting, etc, which may increase or decrease mountain biodiversity ("Data Mining for Global trends in mountain biodiversity", Spehn et al 2005) and, through this, affect slope processes, erosion, water yield and inhabitability.

We can find out whether human activity has affected mountain biodiversity or not by answering the following questions

- Are areas with traditional burning regimes, in combination with grazing, poorer in species of flowering plants, butterflies, and wild ungulates than grazed areas in which burning is not a tradition?
- Do these trends interact with precipitation?
- Is high human population density at high elevations related to the specific loss of woody taxa?
- Is the biological richness of inaccessible microhabitats (topography-caused wildernesses) a measure or good reference of potential bio-diversity of adjacent, transformed land?

These questions can be answered by linking thematic databases for land cover type, population density and climate with regional biodiversity inventories. Comparison of intensively used high-elevation rangeland in regions of contrasting natural biodiversity should illustrate the significance of regional species pools for biodiversity in transformed landscapes A comparison of rangeland biodiversity in geologically young (steep) mountain regions with that in geologically old (smooth) mountain landscapes could reveal interactive influences of landscape roughness and land use on biodiversity.

iv. Assessing mountain biodiversity change under environmental change

With many global mountain biodiversity hotspots increasingly threatened, efforts are required to preserve this unique biota, largely by establishing a system of protected areas on mountains (Koerner and Ohsawa 2005). Relevant variables for conservation biology such as minimum range, viable population size, and connectivity become especially critical in high mountain environments, where range sizes are generally small and where populations are often geographically isolated. In combination with population, genetic, ecological, and phylogeographic data for species of high conservation concern, analysis of such comparative data from different mountain ranges should provide guidelines for critical habitat sizes and minimum coverage of elevational ranges, with the overall task of maximizing the evolutionary potential through phylogenetic diversity and of capturing unique elements of mountain biota

To access change on biodiversity we need to answer following questions

- Which is the minimum altitudinal range required for protected areas in mountain regions?
- What are the minimum habitat size and requirements for long-term viable populations under high mountain conditions and under future climate change?
- Which are the best diversity/area relationships in high mountain environments for conservation purposes?
- What is the relevance of connectivity through gene flow for geographically isolated populations on high mountains?
- Which are suitable indicators and the most likely drivers of biodiversity change in protected areas in mountains?

For conservation planning it will be important to integrate occurrence data across multiple organism groups from different mountain areas, which need to be analyzed in combination with other biotic and abiotic data using information such as in the Global Database of Protected Areas of IUCN and WCMC.

IV. CONVENTION ON BIOLOGICAL DIVERSITY (CBD)

"The CBD's objectives are (1) to conserve biological diversity, (2) to promote the sustainable use of its components, and (3) to achieve fair and equitable sharing of the benefits arising out of the utilization of genetic resources These objectives find expression in the provisions of the CBD, many of which are affected, directly or indirectly, by IPR(intellectual property rights)The relevance of IPRs stems from their role as one of society's principal mechanisms for protecting and enforcing control over information"(Conference Of The Parties To The Convention On Biological Diversity,2012).

A. Intellectual property rights on the conservation and sustainable use of biodiversity.

While related to a number of aspects of biodiversity conservation, IPRs are proving particularly relevant to provisions of the CBD that govern the following four inter related areas

(http://en.wikipedia.org/wiki/Convention_on_Biological_ Diversity)

- Access to and the Fair and Equitable Sharing of Benefits arising from the Utilization of Genetic Resources
- Preservation of and Respect for the Knowledge, Innovations, and Practices of Indigenous and Local Communities
- Transfer of Technology

- Conservation and Sustainable Use of Biological Diversity

An overarching objective of the CBD is encouraging the conservation and sustainable use of the components of biological diversity. This objective encompasses many of the issues raised above, and requires consideration of additional, often indirect, impacts of IPRs on the conservation and sustainable use of biodiversity.

V. REFERENCES

[1] Catherine Monagle(2001) for CIEL and WWF International Biodiversity & Intellectual Property Rights,March 2001

[2] Christian Körner et al(2002) Creative Use of Mountain Biodiversity Databases

[3] Arzberger P et al(2004) An international framework to promote access to data, 19 Mar 2004 303(5665):1777-8

[4] Arzberger P et al (2004) Promoting Access to Public Research Data for Scientific, Economic, and Social Development Data Science Journal 3(29) Volume 3, 29 November 2004 135

[5] Chapman AD(2005) Principles of Data Quality version 1.0. Copenhagen, Denmark: Global Biodiversity Information Facility, 2005

[6] Koerner and Ohsawa(2005) Creative Use of Mountain Biodiversity Databases The Kazbegi Research Agenda of GMBA-DIVERSITAS, Mountain Research and Development Vol 27 No 3 ,Aug 2007

[7] Spehn et al (2005) Data Mining for Global trends in mountain biodiversity,CRC Press,2005

[8] Workshop convened by the Global Mountain Biodiversity Assessment of DIVERSITAS and ICIMOD(2008) Linking Geodata with Biodiversity Information in the Himalayas ICIMOD, Kathmandu, Nepal,15-16 November 2008

[9] Sunita Chaudhary et al (2010) Open Access to, and publication of, Mountain Biodiversity Data of the Hindu Kush Himalayan region ICIMOD, Katmandu, Nepal, 14-18 June 2010

[10] PeterSaundry(2011)Ecosystems_and_Human_Well_Being:V olume1:CurrentState and Trends: Mountain Systems http://www.eoearth.org/article.Updated: 21 September, 2011

[11] Conference Of The Parties To The Convention On Biological Diversity(2012) A review of barriers to the sharing of biodiversity data and information, with recommendations for eliminating them, Eleventh meeting, Hyderabad, India, 8-19 October 2012

[12] HKH Conservation Portal ICIMOD http://www.icimod.org/hkhconservationportal

[13] Convention on Biological Diversity http://en.wikipedia.org/wiki/Convention_on_Biological_ Diversity.