WP 14
23 April 2010

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint Eurostat/UNECE Work Session on Demographic Projections
(28-30 April 2010, Lisbon, Portugal)

Item 6 – Forecasting demographic components: fertility

A stochastic version of the United Nations World Population Prospects: methodological improvements by using Bayesian fertility and mortality projections[1]

Gerhard K. Heilig[2], Thomas Buettner[3], Nan Li[4] and Patrick Gerland[5], United Nations Population Division
Leontine Alkema[6], National University of Singapore
Jennifer Chunn[7] and Adrian E. Raftery[8], University of Washington

1. INTRODUCTION

The United Nations Population Division has prepared population estimates and projections for all countries of the world since the early 1950s. These World Population Prospects (WPP) are now used throughout the whole UN system and by many international organizations, as well as academic researchers. They have become a standard input for development planning, monitoring and global modeling.

Traditionally, the World Population Prospects have included high and low fertility variants, in which the total fertility was assumed to remain 0.5 children above or below the medium variant over most of the projection period. These assumptions are insensitive to the initial level of fertility and to the rate of fertility decline in various countries and may therefore oversimplify or even distort the future range of fertility variation.

Recent research in collaboration with the University of Washington has now enabled the Population Division to calculate probabilistic population projections for all countries of the world. Initial results indicate that the WPP projections might have considerably underestimated the uncertainty in population growth of developing countries- particularly in the Least Developed Countries. The 95% projection intervals of total population growth, as given by the probabilistic projections, are significantly wider in these countries than the high-low variant ranges in the current WPP projections.

[1] This is not a United Nations publication! The views expressed in this paper are those of the authors and do not necessarily reflect the views of the United Nations. Its content has not been formally edited and cleared by the United Nations. The designations employed and the presentation of material in this paper do not imply the expression of any opinion whatsoever on the part of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.
The research of Alkema, Chunn and Raftery was supported by NICHD grant R01 HD054511.
[2] Chief, Population Estimates and Projections Section, United Nations Population Division. e-mail: heilig@un.org
[3] Chief of Population Branch, Deputy Director, United Nations Population Division.
[4] Population Affairs Officers, Population Estimates and Projections Section, United Nations Population Division.
[5] Population Affairs Officers, Population Estimates and Projections Section, United Nations Population Division.
[6] Assistant professor, Department of Statistics and Applied Probability, National University of Singapore.
[7] Graduate Student, University of Washington.
[8] Blumstein-Jordan Professor of Statistics and Sociology at the University of Washington.

2. PREVIOUS STUDIES

Traditionally, population projections have used variants or scenarios in the projection of fertility, mortality and migration to indicate uncertainty or illustrate policy options. The variants are usually based on plausible alternatives, such as the "half-child" rule for high and low fertility projections in the United Nations *World Population Prospects*. This pragmatic approach has the advantage that results can be easily communicated to policy makers and planners. However, it has the disadvantage that variants do not indicate anything about their probability [1]. They are just plausible alternatives. Demographers have therefore developed probabilistic population projections where the range of uncertainty is represented by projection intervals. Probabilistic projections come in various flavors [2]: Some authors have used historical forecast errors [3,4]; others have based their projections on expert opinion [5,6] or used time series analysis to project future population parameters [7-9]. Existing probabilistic population projections so far have been limited in their geographical scope either to macro-regions at the global scale [10-12], a specific region such as Europe [13-15], or few countries for in-depth case studies [16-21]. So far, no attempt has been made to produce national-level probabilistic population projections for all countries of the world [22].

3. OUR APPROACH TO PROBABILISTIC PROJECTIONS

We have used a Bayesian statistical approach to project total fertility and life expectancy at birth between 2005 and 2050, based on estimates between 1950 and 2005 [23-25]. The fertility projection model consists of two components. The first component models the fertility decline from high or medium levels of fertility to a turnaround point close to replacement level fertility. The second component models changes in total fertility after the turnaround point. These changes are random fluctuations around and/or recovery towards replacement level fertility, which are modeled using a first order autoregressive time series model, an AR(1) model, with its mean fixed at approximate replacement-level fertility (2.1). In the projections of life expectancy at birth, expected gains in life expectancy are modeled as a function of its level, with random distortions added to it.

The projection methodology builds on the current methodology used in the World Population Prospects. Expected 5-year decrements in total fertility are modeled as a function of its level using a bi-logistic function[26]. Similarly, expected 5-year gains in life expectancy at birth are modeled as a function of its level, using a bi-logistic function [26]. In the Bayesian projection model, instead of using a given set of parameters to find the fertility decrements and gains in life expectancy in a country, a Bayesian hierarchical model [27,28] is used to estimate these parameters for each country. Each of the unknown parameters of the bi-logistic functions is drawn from a probability distribution that represents the range of outcomes of that parameter across all countries. For a specific country, the posterior distribution of its parameters is determined by the world-level experience for all countries combined, as well as the observed declines in that country..

The Bayesian projection models yield a large number of trajectories of future total fertility and life expectancy, for all countries without generalized HIV/AIDS epidemic (i.e., less than 1% HIV prevalence in the general adult population age 15-49) and with a total population larger than 100,000 in 2009. We have used a male-dominant projection of life expectancy, and derived female life expectancies by assuming the same sex differentials as projected deterministically by the United Nations *World Population Prospects*. From the large number of trajectories (35,000), 1,000 trajectories were sampled to be used in standard cohort-component population projections [29,30]. To convert the projections of total fertility into the age-specific fertility rates as required by the cohort component projection, percentages of age-specific fertility rates were taken from the *2008 Revision* of the *World Population Prospects* and applied throughout the projection period. Projections of life expectancy at birth were converted into age- and sex-specific mortality rates and, subsequently, into survivor ratios which were then used in the cohort-component projections. For the historical periods, we used a standard Lee-Carter model [7]. For the future, we applied a modified Lee-Carter method, in which the time indicator of a Lee-Carter model was not projected by a random walk (as is the standard method), but was fitted to the trajectories of life expectancy at birth, as projected by the Bayesian projection model.

All other input data, including base population by age and sex as well as projections of net-migration, were identical to the *2008 Revision* of the *World Population Prospects* [31]. Using the 1,000 sets of age-specific fertility rates and sex-specific survivor ratios, 1000 cohort component projections were calculated for the period 2010 to 2050 for all 196 countries with a population of more than 100,000 in 2009. From these trajectories the median of total population and a 95% projection interval was calculated for each country. These first results of our probabilistic population projections were compared to the high, low and medium variant of the *2008 Revision* of the *World Population Prospects* [31] (see figures with results in Appendix).

4. RESULTS

We have calculated probabilistic population projections for 196 countries of the world with a population of more than 100,000 (Atotal population projections, including 95% projection intervals and median) and compared these projections with the deterministic projections as provided by the *2008 Revision* of the *World Population Prospects* [31]. This work has yielded four important results:

(1) While probabilistic projections are available for particular countries or regions this is the first time that probabilistic projections were calculated in a systematic and consistent way for almost all countries of the world – including 151 developing countries. We now can calculate projection intervals that are based on probability, rather then projection intervals that are based on high-low scenarios, as in the deterministic *World Population Prospects*. This is of particular relevance for developing countries, where large differences in the availability and quality of the empirical input data for the initial estimates are common. With the probabilistic projections these varying uncertainties in estimates are used to determine the probability of the projection intervals. The global coverage also allows us to analyze differences between probabilistic and deterministic projections for countries having very different initial levels of fertility and mortality.

(2) We found that, in more than 80% of the countries, probabilistic medians in projected total population differ less than 10% from the medium variant of the *World Population Prospects*. Figure 1 illustrates the high correlation between the probabilistic median and the (deterministic) medium variant population projection for the year 2050. However, we also found that in less than 20% of the countries of the world probabilistic projections resulted in substantially higher or lower median projections than were previously projected in the (deterministic) *World Population Prospects*. This concerns primarily countries where the level of fertility is still rather high or where fertility declines have been rapid and recent.

(3) We could confirm that in about 60% of the countries, the high-low scenarios of the (deterministic) *World Population Prospects* lie within the 95% confidence interval of the probabilistic population projections. This indicates, for the first time, how the United Nations projection scenarios are related to the uncertainty determined by a probabilistic projection.

(4) We also found that the range of uncertainty as measured by a 95% projection interval in our probabilistic population projections for 2050 was much larger for some countries than the high-low variant range in the deterministic projections. Figure 2 plots the ratio of the 95% projection interval to the deterministic high-low range in population projections for 2050 against the level of total fertility in the five-year interval of 2000-2005. The figure reveals that in countries with high fertility in 2000-2005 the probabilistic projection resulted in a much wider 95% projection interval than the high-low variant range (i.e., +/- 0.5 child) in the *World Population Prospects*. As indicated by Map 1, these countries are mainly located in Africa and often belong to the group of Least Developed Countries (see appendix 1). These results seem to indicate that the range of future outcomes as given by the high-low variants in the deterministic *World Population Prospects* is narrower than the likely range of possible future scenarios in population projections for countries, which are still at the early stages of their fertility decline.

5. DISCUSSION

Uncertainty in population projections results from three major sources[32]: the accuracy of the input data, the methods and assumptions for projecting future trends in fertility, mortality and migration, and the intrinsic unpredictability of all socio-economic events, which include demographic events such as migration.

The first source of uncertainty is identical between our probabilistic and deterministic projections. Both use historical *estimates* of population by age and sex, total fertility, life expectancy at birth and net-migration that had been prepared for the *2008 Revision* of the *World Population Prospects*. These estimates are based on a large number of empirical data sources, including censuses, population and vital registers as well as surveys, such as the Demographic and Health Surveys (DHS) the Multi-Cluster Indicator Surveys (MICS) and others. Typically, these empirical data sources do not provide unambiguous historical time series of demographic indicators, but widely ranging “data clouds” – particularly in developing countries. If the Bayesian projection model would be based on these raw, unadjusted data sets, the projection intervals would certainly be much wider. However, not all of the variation in these empirical data can be used in assessing uncertainty. Some of the empirical information is obviously wrong or deficient and can be corrected or adjusted by established demographic methods – such as the age-heaping in population counts or the underreporting of birth and children under the age of five. The Population Division is using various methods to check consistency between data sources, calculate intercensal growth balances, compare countries within a region and apply adjustments, if necessary [33]. This estimation procedure results in relatively smooth historical time series of total fertility and life expectancy at birth in five-

year intervals, which certainly hide some of the actual annual fluctuations and uncertainties in demographic indicators, but also corrects unrealistic variation due to errors and data deficiencies. The projections based on the UN estimates can be viewed validly as probabilistic projections of future UN estimates.

In the Bayesian projection model, the projections of future trends in fertility and mortality are based on the historical trends of fertility and mortality declines in all countries in the world. The projected variance thus results from the variance in the pace of observed declines of fertility and mortality in all countries. This method of taking into account uncertainty seems more appropriate than assuming a single high or low variant of fertility decline for all countries, as in the (deterministic) *World Population Prospects*. Therefore, we likely improve the second source of uncertainty.

The third source of uncertainty is the fact that human behavior is intentional and ultimately unpredictable. While the past decline of fertility in most countries strongly indicates a general trend towards low fertility, it is possible that some populations might not follow these trends. Demographers have discussed the phenomenon of "stalled fertility decline" [34,35]. The universal trends towards higher life expectancy might also slow down or even reverse due to massive increases in unhealthy and risky life styles (obesity, alcohol misuse), emergence of devastating pandemics, or widespread wars. Migration, in particular, will always include a component of inherent unpredictability, because it can be induced or stopped by actions of governments and by natural conditions. No one can exclude the possibility of massive displacement of people due to conflict, economic collapse, or natural disaster[36]. We have partially accounted for this source of uncertainty by adding random distortion terms to future trajectories. The variance of these terms is estimated based on observed past distortions in all countries.

6. CONCLUSION

While population projections will always include several elements of uncertainty, our probabilistic projections allow us to quantify those aspects of uncertainty which can be derived from countries' past experience. The Bayesian Hierarchical Model is data-driven, since it uses the UN estimates which are based on empirical data. It does not depend on expert opinion to determine uncertainties and is fully reproducible.

Our probabilistic projections show that the uncertainty in future total population trends in countries with high fertility is probably larger than indicated by the range between high and low variants in the United Nations *World Population Prospects*.

Figure 1: Total Population in 2050: Probabilistic Median versus Medium Variant WPP

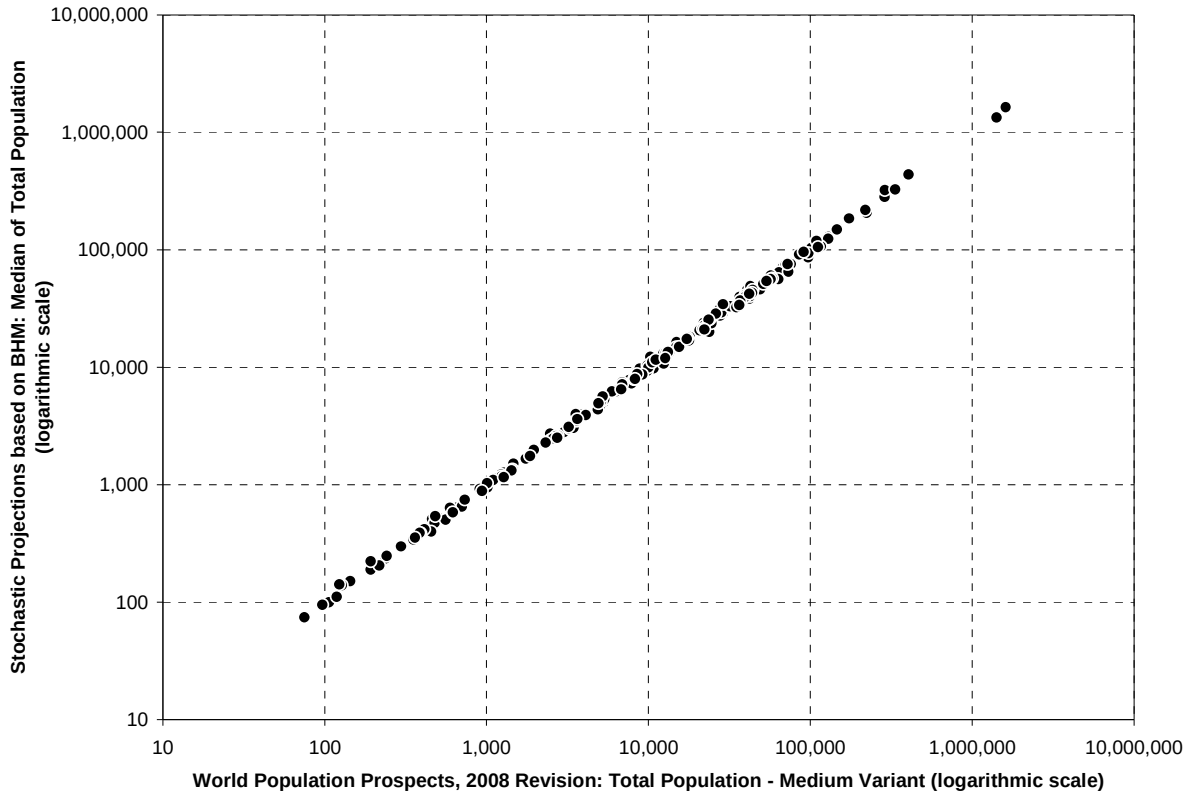
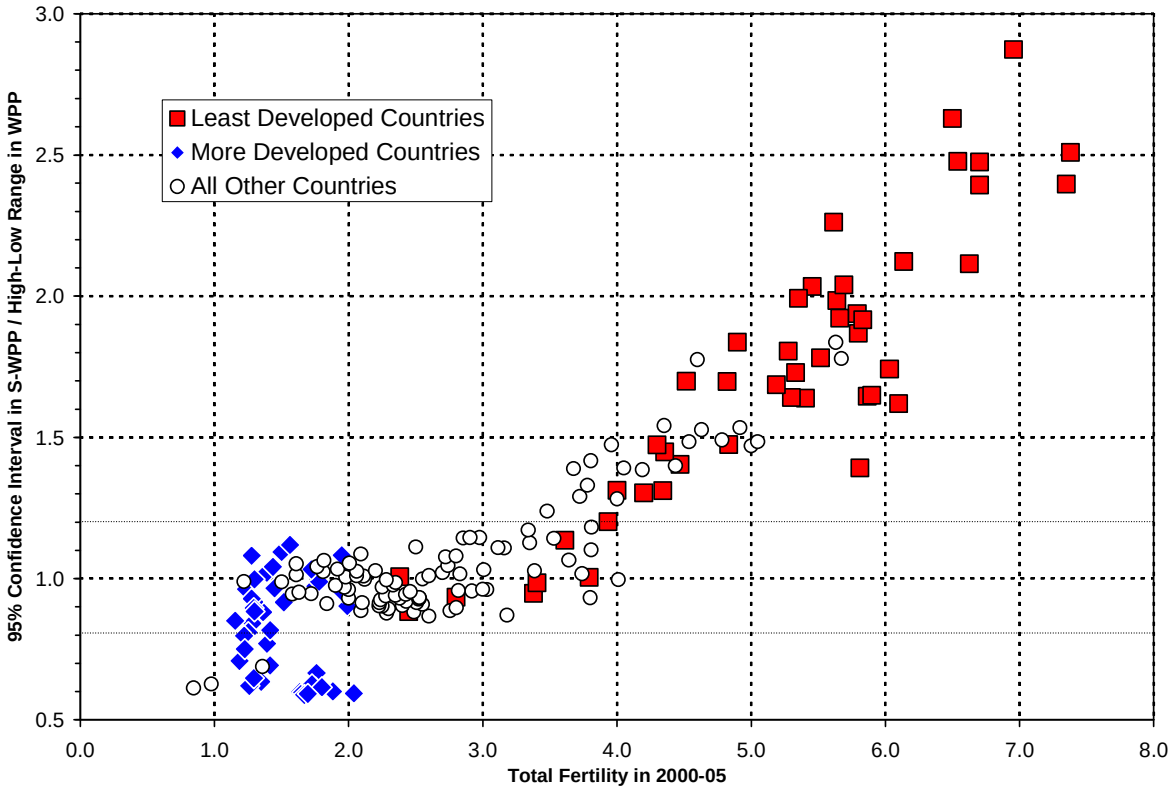
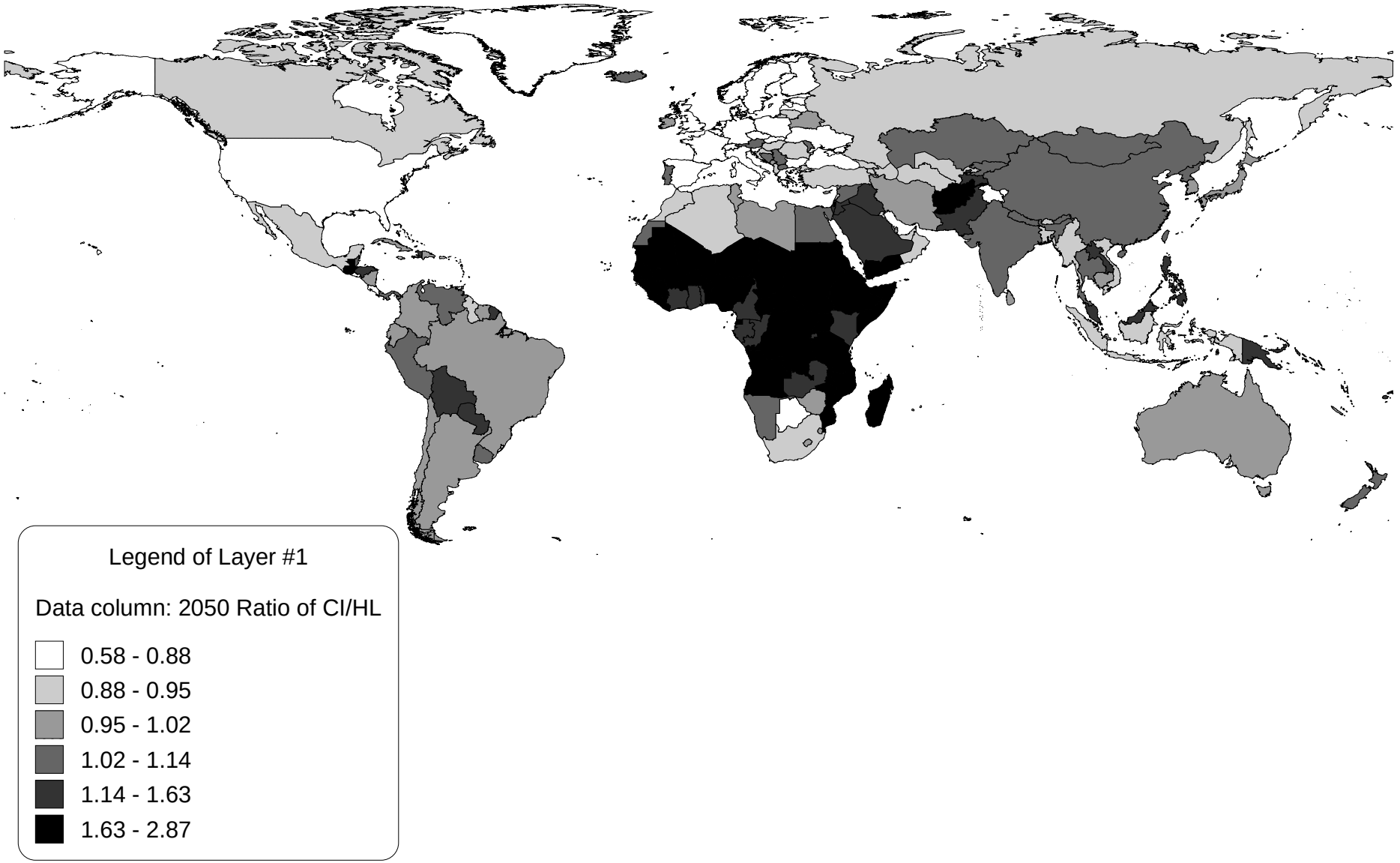


Figure 2: Ratio of probabilistic projection interval and high-low range in WPP versus total fertility in 2005-10



Map 1: Ratio of probabilistic projection interval and high-low range in WPP versus total fertility in 2005-10



References

1 Keilman, N., Pham, D. Q. & Hetland, A. Why population forecasts should be probabilistic - illustrated by the case of Norway. *Demographic Research* **6**, 409-454, doi:10.4054/DemRes.2002.6.15 (2002).

2 Booth, H. Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting* **22**, 547-581, doi:10.1016/j.ijforecast.2006.04.001 (2006).

3 Keyfitz, N. The Limits of Population Forecasting. *Population and Development Review* **7**, 579-593 (1981).

4 Stoto, M. A. The Accuracy of Population Projections. *Journal of the American Statistical Association* **78**, 13-20 (1983).

5 Lutz, W., Sanderson, W. & Scherbov, S. in *The future population of the world : what can we assume today?* (ed Wolfgang Lutz) 397-428 (Earthscan Publications, 1996).

6 Lutz, W., Sanderson, W. C. & Scherbov, S. in *Frontiers of population forecasting, A supplement to Population and Development Review* Vol. 24 eds W. Lutz, J.W. Vaupel, & D.A. Ahlburg) 139-155 (Population Council, 1998).

7 Lee, R. D. & Carter, L. R. Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association* **87**, 659-671 (1992).

8 Lee, R. D. Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting* **9**, 187-202, doi:10.1016/0169-2070(93)90004-7 (1993).

9 Lee, R. & Miller, T. Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. *Demography* **38**, 537-549, doi:10.1353/dem.2001.0036 (2001).

10 Lutz, W., Sanderson, W. & Scherbov, S. Doubling of world population unlikely. *Nature* **387**, 803-805, doi:10.1038/42935 (1997).

11 Lutz, W., Sanderson, W. & Scherbov, S. The end of world population growth. *Nature* **412**, 543-545, doi:10.1038/35087589 (2001).

12 Lutz, W., Sanderson, W. & Scherbov, S. The coming acceleration of global population ageing. *Nature* **451**, 716-719, doi:10.1038/nature06516 (2008).

13 Alders, M., Keilman, N. & Cruijsen, H. Assumptions for long-term stochastic population forecasts in 18 European countries. *Hypothèses de projections stochastiques à long terme des populations de 18 pays européens.* **23**, 33-69, doi:10.1007/s10680-006-9104-4 (2007).

14 Alho, J. *et al.* New forecast: Population decline postponed in Europe. *Statistical Journal of the UN Economic Commission for Europe* **23**, 1-10 (2006).

15 Alho, J. M., Jensen, S. E. H. & Lassila, J. *Uncertain Demographics and Fiscal Sustainability*. (Cambridge University Press, 2008).

16 Lee, R. D. & Tuljapurkar, S. Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low. *Journal of the American Statistical Association* **89**, 1175-1189 (1994).

17 Alho, J. M. A stochastic forecast of the population of Finland. (Statistics Finland, Helsinki, 1998).

18 Lutz, W. & Scherbov, S. An expert-based framework for probabilistic national population projections: The example of Austria. *European Journal of Population-Revue Europeenne De Demographie* **14**, 1-17 (1998).

19 Lutz, W. & Scherbov, S. Probabilistic Population Projections for India with Explicit Consideration of the Education-Fertility Link. *International Statistical Review* **72**, 81-92 (2004).

20 Hyndman, R. J. & Booth, H. Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting* **24**, 323-342, doi:10.1016/j.ijforecast.2008.02.009 (2008).

21 Booth, H., Pennec, S. & Hyndman, R. Stochastic population forecasting using functional data methods: the case of france. (XXVI IUSSP International Population Conference, Marrakech, Morocco, 2009).

22 Bongaarts, J. & Bulatao, R. A. *Beyond Six Billion : Forecasting the World's Population. Panel on Population Projections, National Research Council (U.S.).* (National Academy Press, 2000).

23 Raftery, A. *et al.* in *Expert Group Meeting on Recent and Future Trends in Fertility* (eds Department of Economic and Social Affairs & Population Division) (United Nations, New York, NY, 2009).

24 Alkema, L. *et al.* Probabilistic Projections of the Total Fertility Rate for All Countries. Report No. Working Paper no. 97, (University of Washington, Seattle, WA, 2010).

25 Chunn, J., Raftery, A. & Gerland, P. Bayesian Probabilistic Projections of Mortality. (Annual Meeting of the Population Association of America, Dallas, TX, 2010).

26 United Nations. *World Population Prospects: The 2004 Revision, Volume III: Analytical Report*. (Department of Economic and Social Affairs, Population Division, 2006).

27 Lindley, D. V. & Smith, A. F. M. Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society. Series B (Methodological)* **34**, 1-41 (1972).

28 Gelman, A. *Bayesian data analysis*. xxv, 668 (Chapman & Hall/CRC, 2004).

29 United Nations. *Manual III: Methods for population projections by sex and age*. Vol. 25 - Population Studies (Dept. of Economic and Social Affairs, Population Division, 1956).

30 United Nations. *The United Nations Population Projection Computer Program : A User's Manual*. (Department of International Economicand Social Affairs, Population Division, 1989).

31 United Nations. *World Population Prospects: The 2008 Revision. CD-ROM Edition - Extended Dataset in Excel and ASCII formats*. (Dept. of Economic and Social Affairs, Population Division, 2009).

32 Lee, R. D. Probabilistic Approaches to Population Forecasting. *Population and Development Review* **24 (Supplement: Frontiers of Population Forecasting)**, 156-190 (1998).

33 Heilig, G. K. *et al.* The 2008 Revision of the United Nations World Population Prospects: Challenges in Estimating and Projecting the World's Population. (XXVI IUSSP International Population Conference, Marrakech, Morocco, 2009).

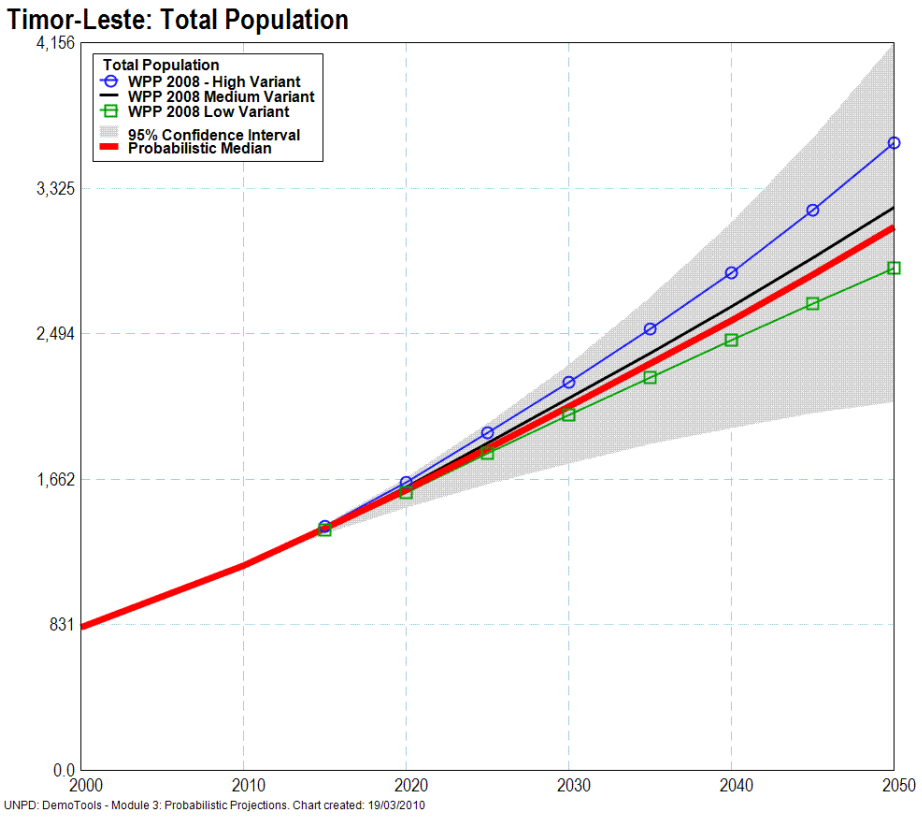
34 Bongaarts, J. The causes of stalling fertility transitions. *Studies in Family Planning* **37**, 1-16, doi:10.1111/j.1728-4465.2006.00079.x (2006).

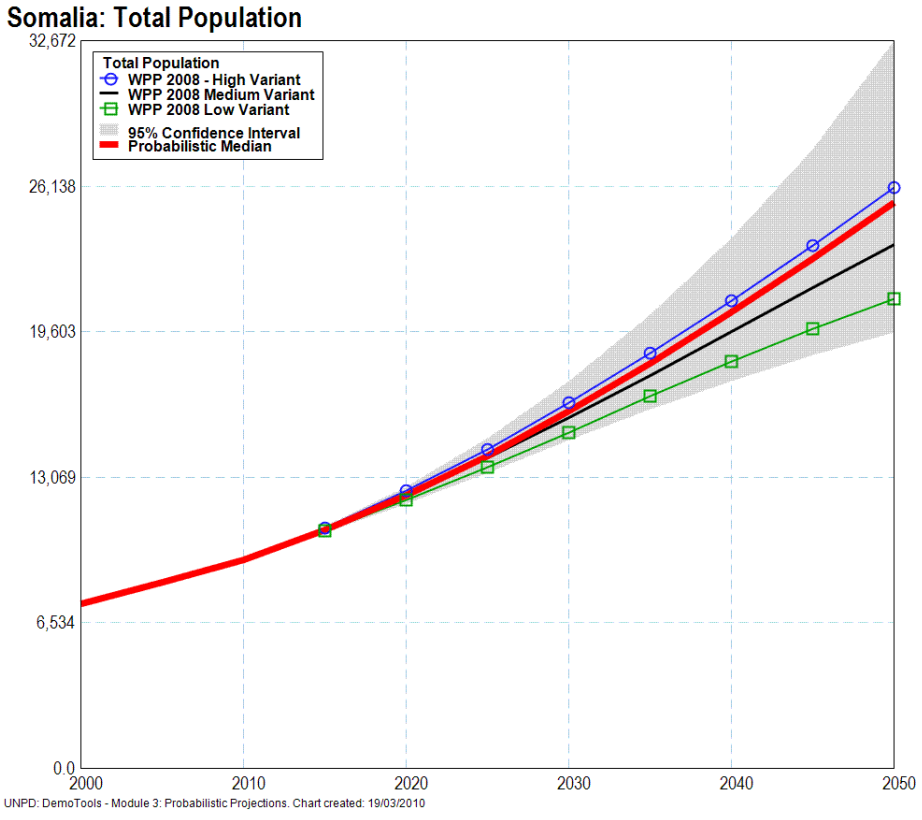
35 Bongaarts, J. Fertility transitions in developing countries: Progress or stagnation? *Studies in Family Planning* **39**, 105-110, doi:10.1111/j.1728-4465.2008.00157.x (2008).

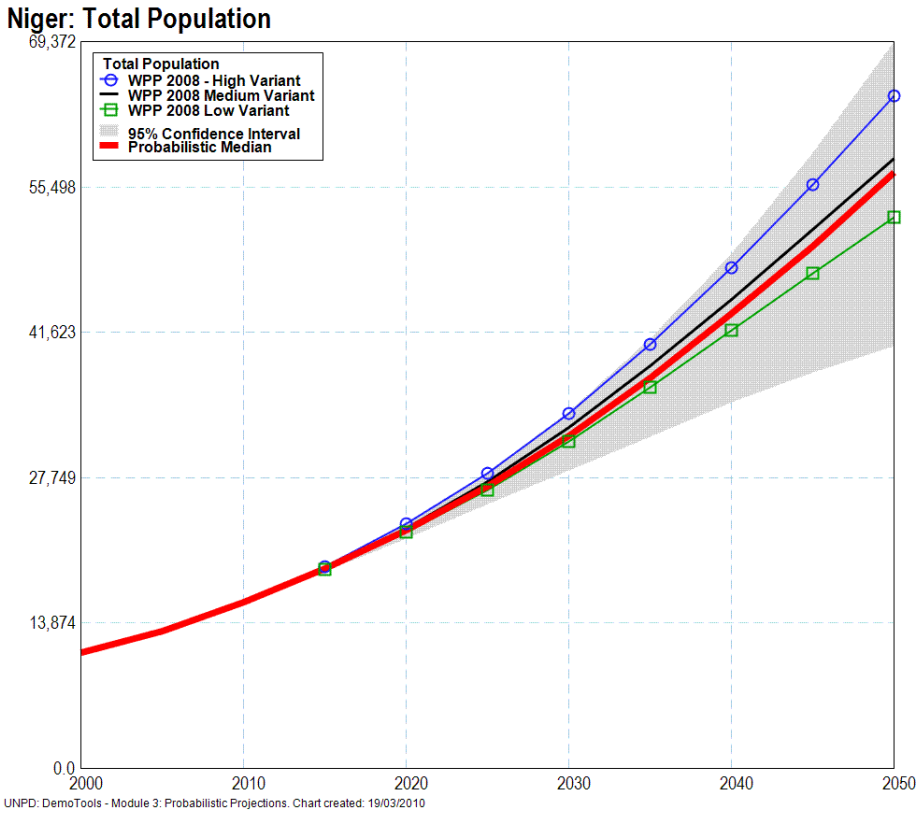
36 Smil, V. The Next 50 Years: Fatal Discontinuities. *Population and Development Review* **31**, 201-236, doi:10.1111/j.1728-4457.2005.00063.x (2005).

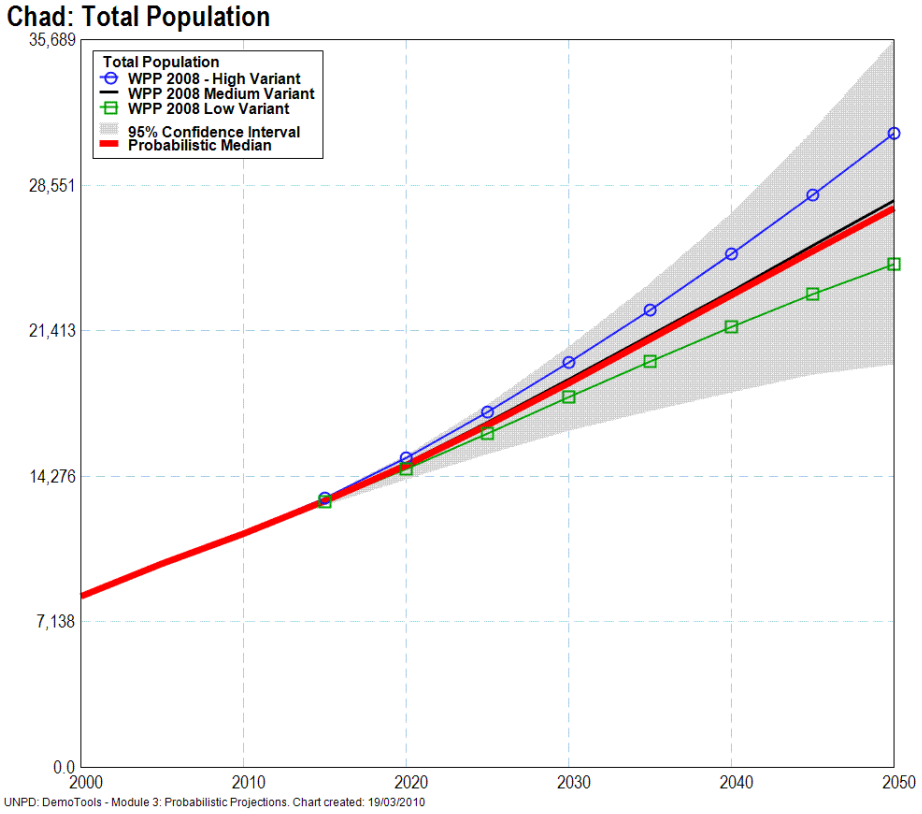
Appendix 1

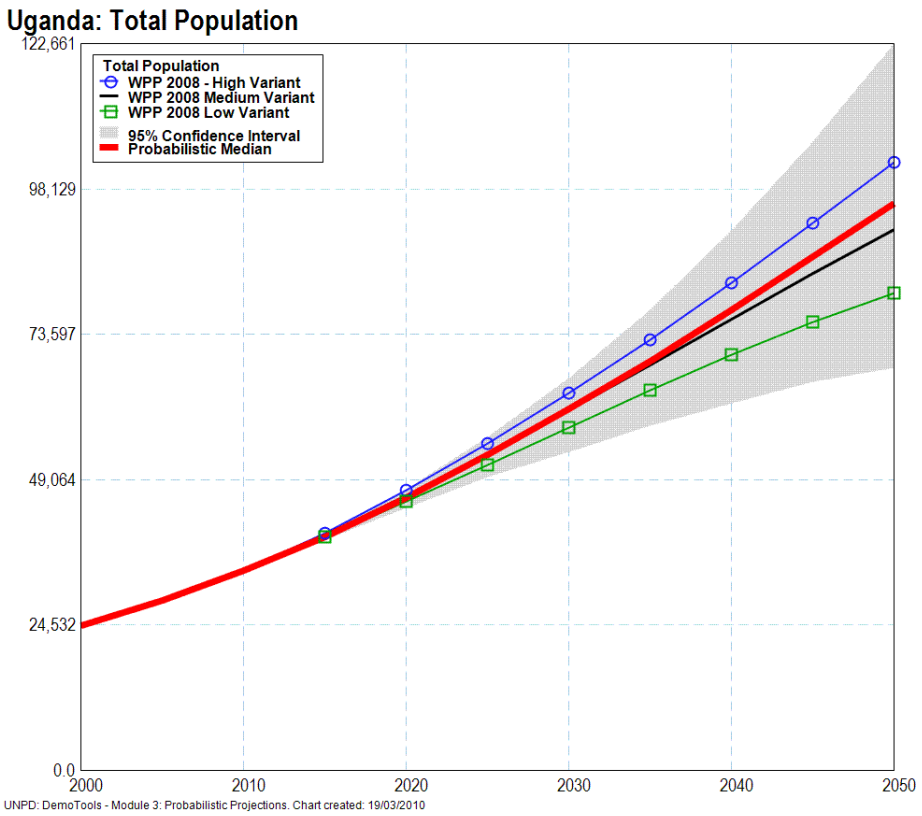
Total population of selected Least Developed Countries, 2000-2050:
Probabilistic projections and World Population Prospects (2008 Revision)
(Projection Base Year: 2005)

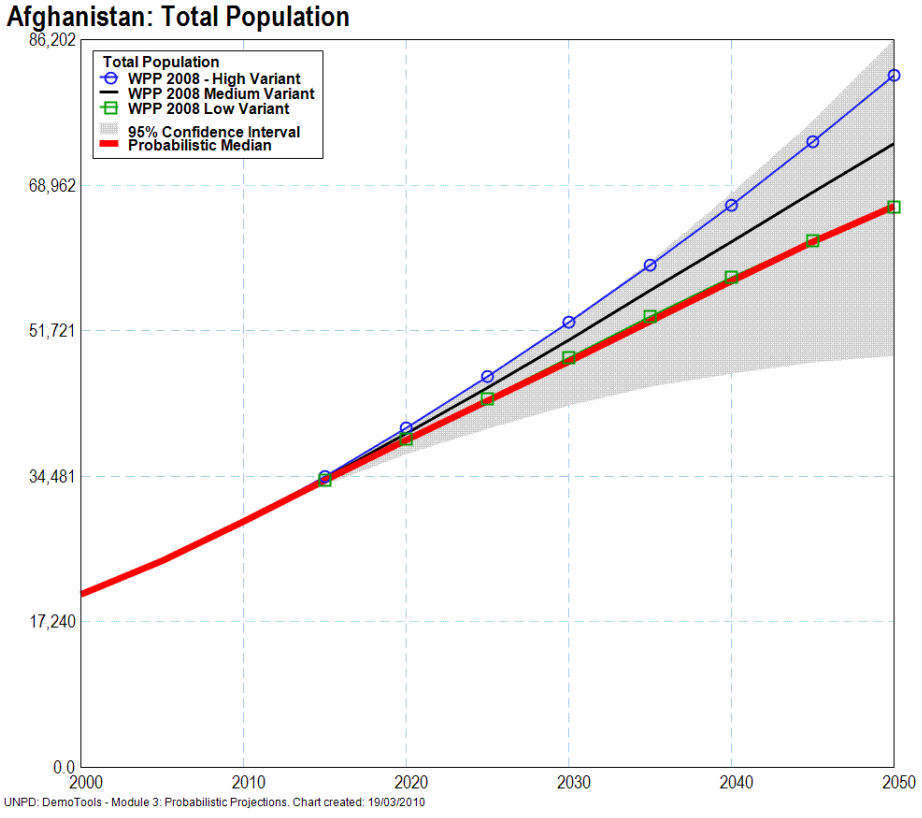

Timor-Leste: Total Population
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
4,156
3,325
2,494
1,662
831
0.0
2000
2010
2020
2030
2040
2050
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010


Somalia: Total Population
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
32,672
26,138
19,603
13,069
6,534
0.0
2000
2010
2020
2030
2040
2050
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010


Niger: Total Population
69,372
55,498
41,623
27,749
13,874
0.0
2000
2010
2020
2030
2040
2050
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010


Chad: Total Population
35,689
28,551
21,413
14,276
7,138
0.0
2000
2010
2020
2030
2040
2050
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010

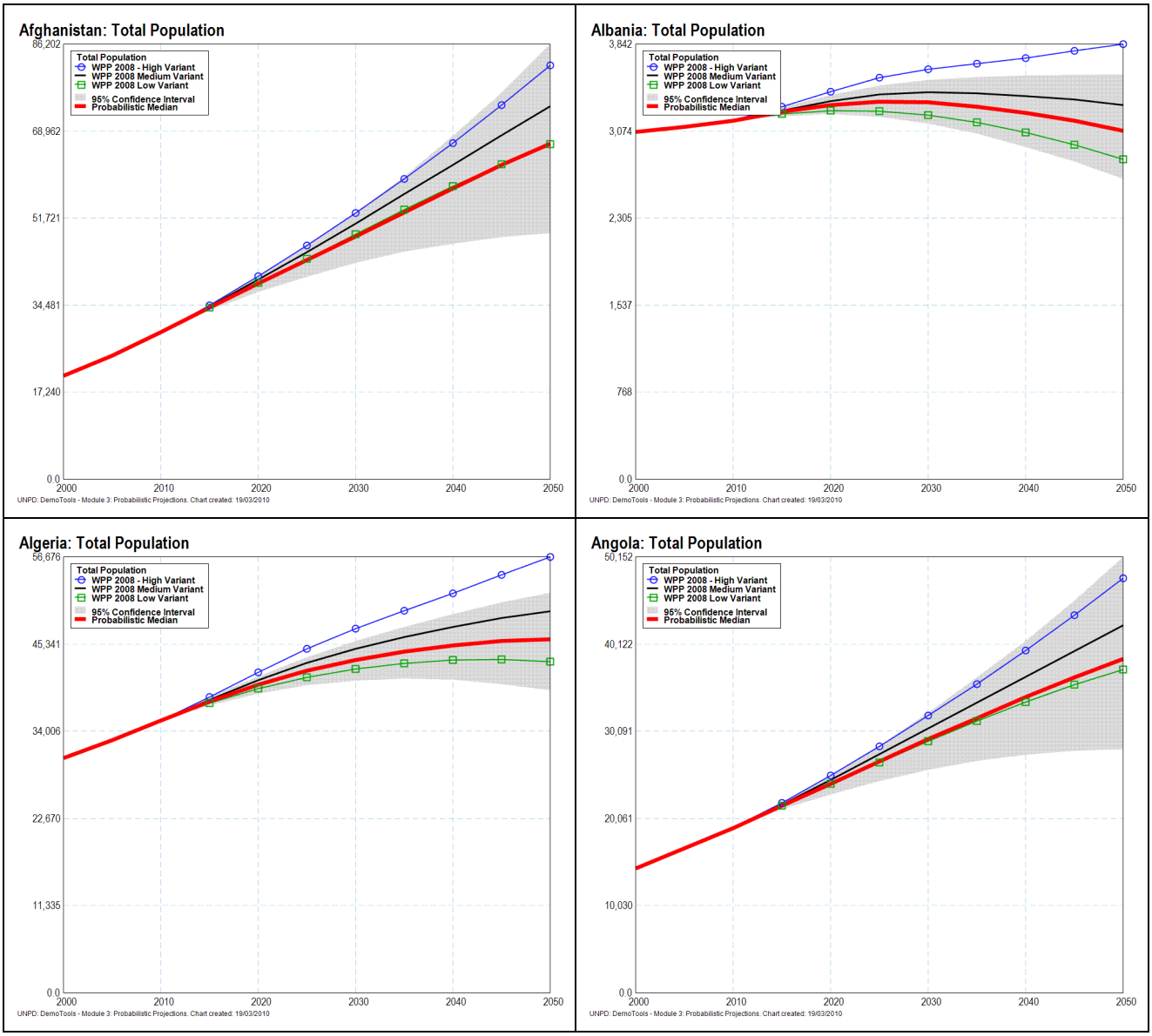

Uganda: Total Population
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
122,661
98,129
73,597
49,064
24,532
0.0
2000
2010
2020
2030
2040
2050
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010

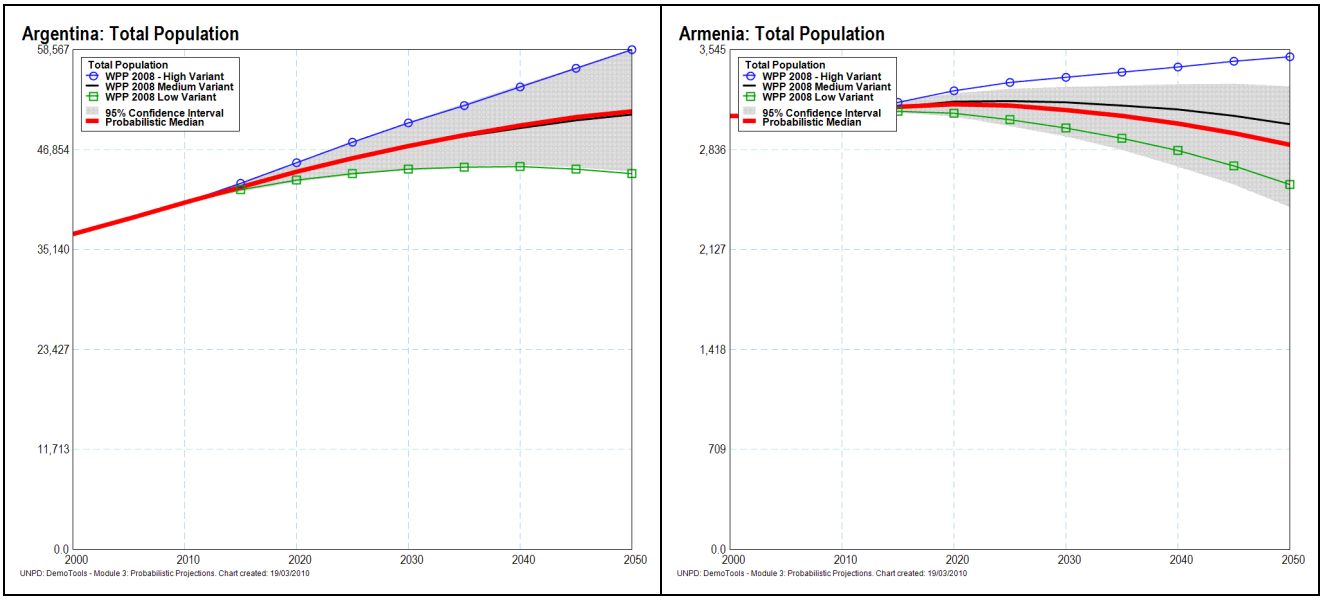

Afghanistan: Total Population
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
86,202
68,962
51,721
34,481
17,240
0.0
2000
2010
2020
2030
2040
2050
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010

Appendix 2

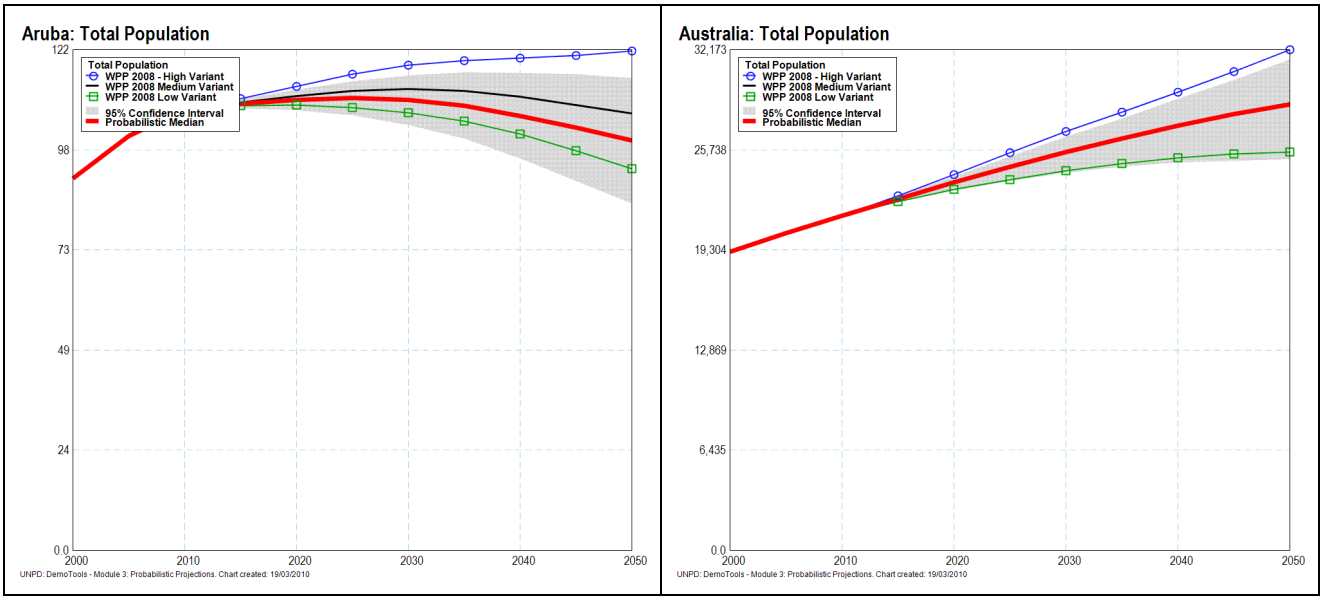
Total population of selected countries, 2000-2050
(Projection Base Year: 2005)

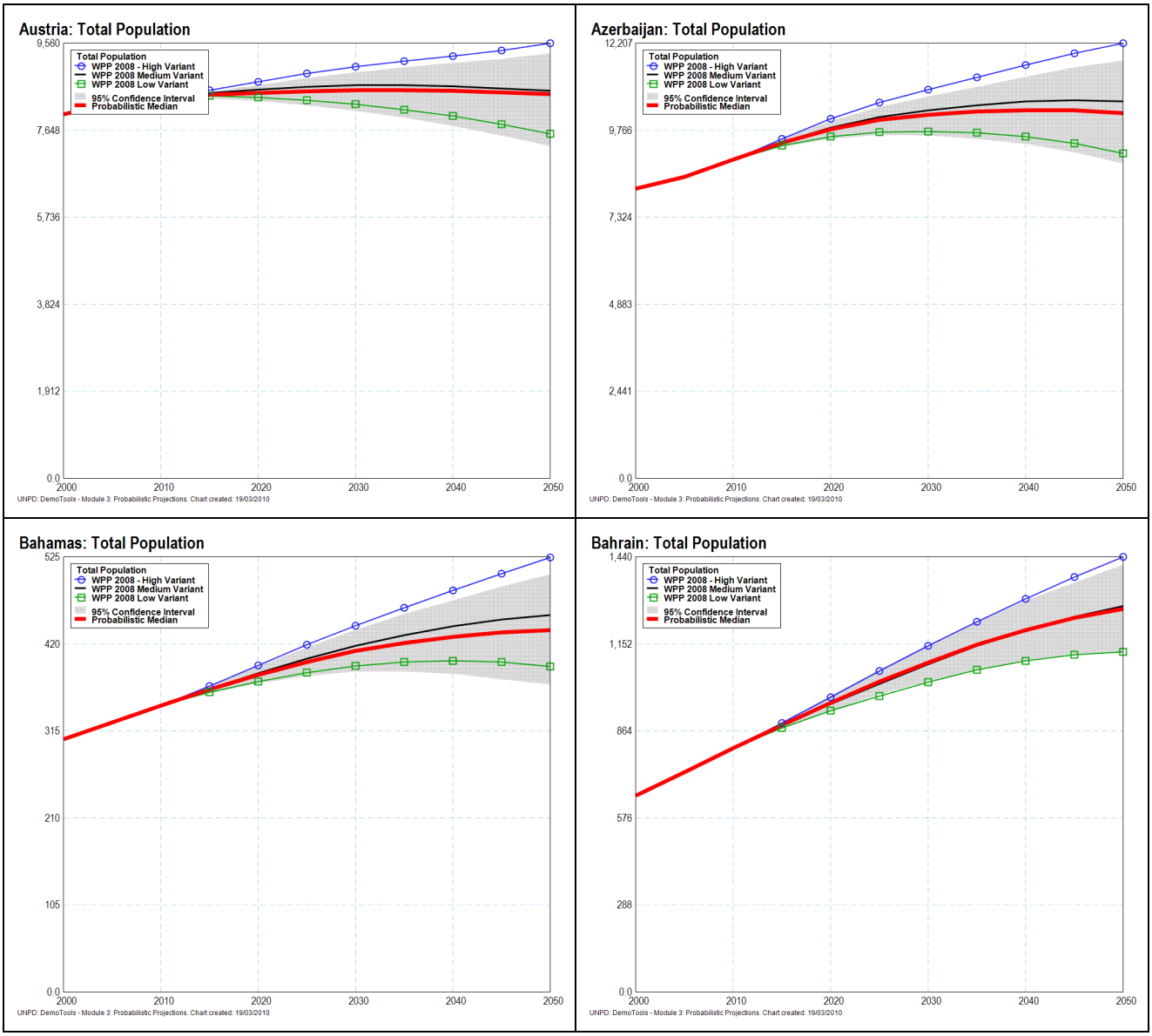
These probabilistic population projections have been computed using stochastic projections of fertility and mortality, generated by a Bayesian Hierarchical Model (Raftery / Alkema, et. al.). The projections are available for all countries of the world with a population of more than 100,000 in 2008. They have been plotted together with the high, medium and low variants of the *2008 revision* of the *World Population Prospects*.


Afghanistan: Total Population
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
86,202
68,962
51,721
34,481
17,240
0.0
2000
2010
2020
2030
2040
2050
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010
Albania: Total Population
3,842
3,074
2,305
1,537
768
0.0
Algeria: Total Population
56,676
45,341
34,006
22,670
11,335
Angola: Total Population
50,152
40,122
30,091
20,061
10,030

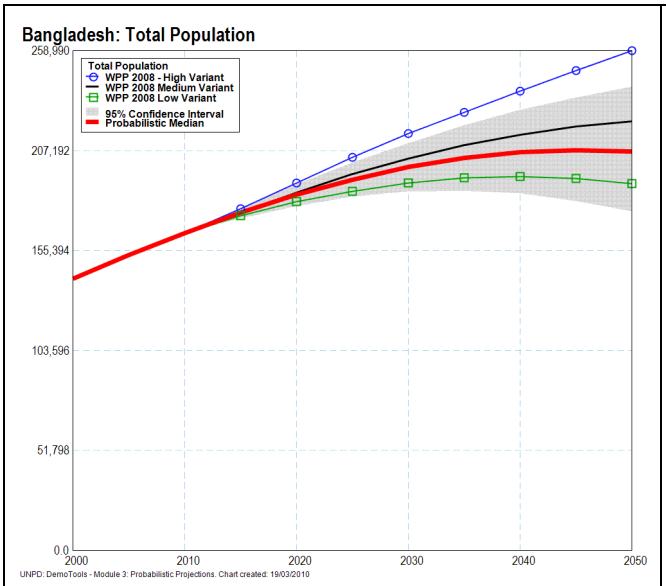


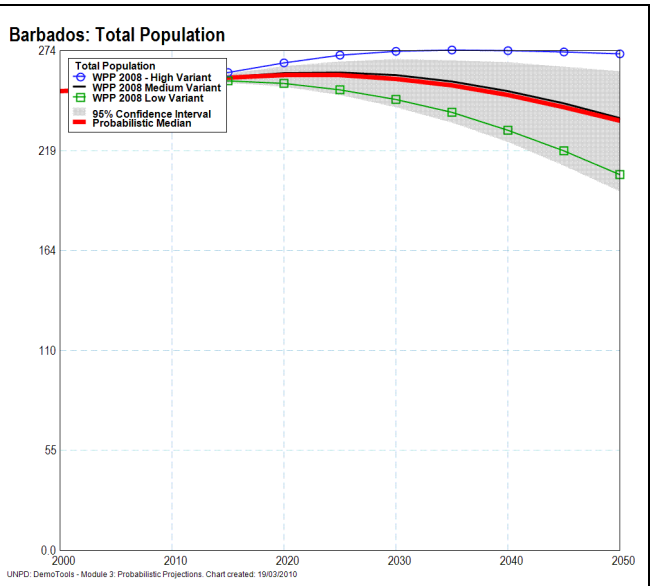
(All population numbers are in thousands)

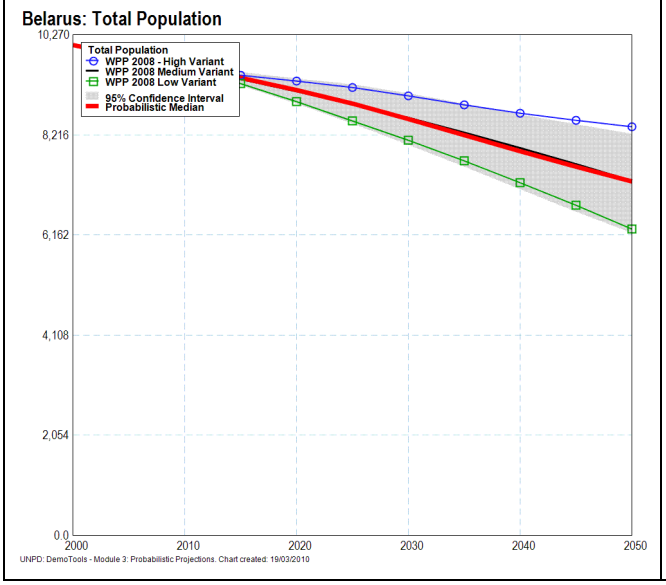


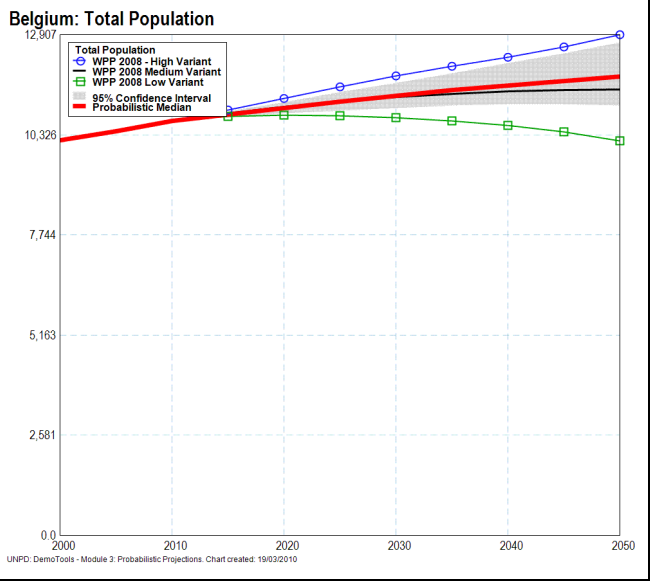


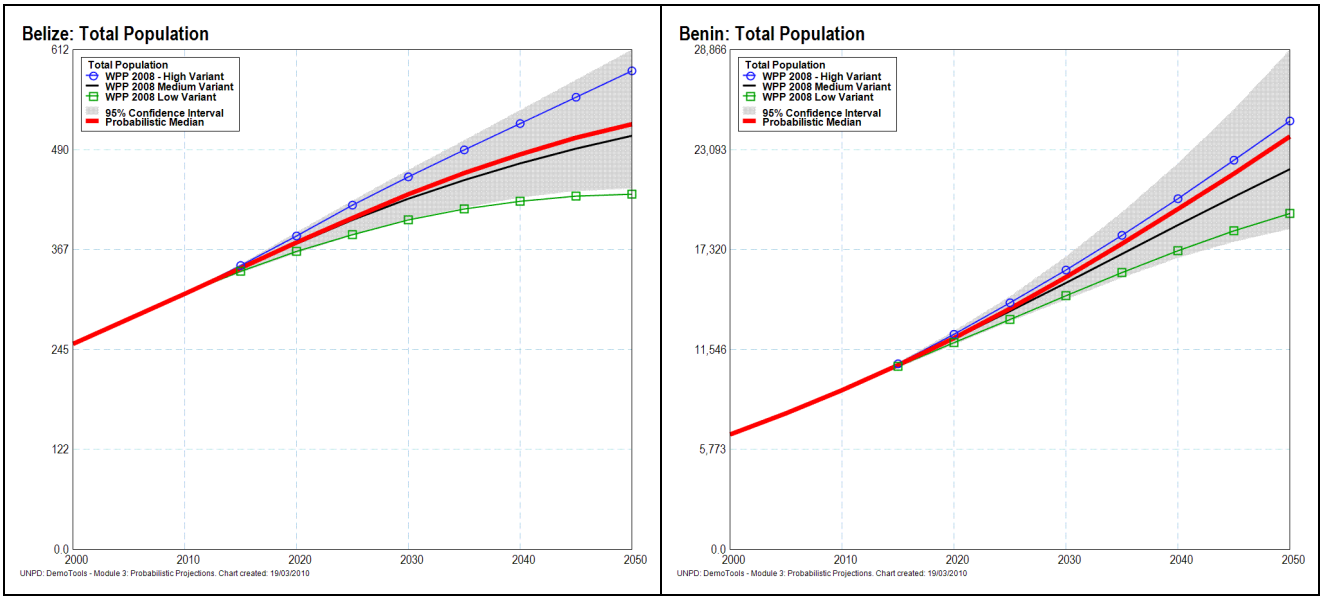
(All population numbers are in thousands)


Bangladesh: Total Population
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
258,990
207,192
155,394
103,596
51,798
0.0
2000
2010
2020
2030
2040
2050
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010


Barbados: Total Population
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
274
219
164
110
55
0.0
2000
2010
2020
2030
2040
2050
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010


Belarus: Total Population
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
10,270
8,216
6,162
4,108
2,054
0.0
2000
2010
2020
2030
2040
2050
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010


Belgium: Total Population
Total Population
WPP 2008 - High Variant
WPP 2008 Medium Variant
WPP 2008 Low Variant
95% Confidence Interval
Probabilistic Median
12,907
10,326
7,744
5,163
2,581
0.0
2000
2010
2020
2030
2040
2050
UNPD: DemoTools - Module 3: Probabilistic Projections. Chart created: 19/03/2010



(All population numbers are in thousands)