

Верификация данных наблюдений о температуре поверхности моря

Н.Б. Захарова

Институт вычислительной математики РАН, Москва, 119333, Россия

E-mail: zakharova_nb@inm.ras.ru

Настоящая работа посвящена обработке гидрофизических данных наблюдений. В работе показана необходимость проводить дополнительную верификацию данных наблюдений, прежде чем использовать их для решения различных задач гидротермодинамики, т.к. даже после калибровки и валидации в центрах, предоставляющих информацию, в данных могут оставаться ошибки.

Необходимость проведения дополнительной верификации данных обоснована на примере оперативных данных наблюдений о температуре поверхности Балтийского моря, получаемых со спутников. В работе проведен анализ данных о температуре поверхности моря за 2015 г. на основе физических свойств исследуемой среды, с выявлением ошибок в данных наблюдений. Значения некоторых полей температуры сильно отличались от полей в соседние моменты времени, что привело к необходимости проведения дополнительных проверок данных.

Описан метод, реализованный для проведения дополнительной верификации получаемых данных, основанный на статистических подходах, которые позволяют установить общие особенности, характерные для всего набора реализаций поля температур. Представлены результаты работы указанных процедур. Численные эксперименты проводятся с использованием статистических данных о температуре поверхности моря – средне-суточных и среднемесячных за 25–27 лет. Реализованный метод позволяет исключить поля данных, значения которых не физичны, т.е. решает поставленную задачу дополнительной верификации данных наблюдений.

Ключевые слова: математическое моделирование, данные наблюдений, верификация данных, температура поверхности моря, обработка данных

Одобрена к печати: 10.05.2016

DOI: 10.21046/2070-7401-2016-13-3-106-113

Данные наблюдений

В данной работе речь идет о верификации гидрофизических данных наблюдений на основе данных о температуре поверхности моря (ТПМ). В качестве примера использовались данные европейского проекта Коперникус (<http://www.copernicus.eu/>) – это международная программа по наблюдению за Землей, которая дает много возможностей для изучения, исследования, мониторинга окружающей среды. Программа включает как международные проекты и разработки ученых из разных стран по моделированию состояния окружающей среды, так и большой объем данных наблюдений за состоянием отдельных сред, полученный объединенными силами стран-участников. Проект обладает большим потенциалом и несет в себе много полезного знания и инструментария для исследователей в области моделирования и мониторинга – как земной поверхности, так и атмосферы и морской среды. Далее при упоминании проекта будем говорить о Сервисе по мониторингу и моделированию состояния морских сред (<http://marine.copernicus.eu/>), поддерживаемом проектом Коперникус, который использовался в настоящей работе. Данные, представленные на портале проекта, являются объединением оперативных данных наблюдений и статистических полей различных параметров, посчитанных вычислительными центрами. В проекте ведется постоянная работа по наполнению портала данными с современных систем наблюдения за состоянием окружающей среды. Представленные на указанном

ресурсе данные обновляются регулярно и представляют собой довольно точную информацию об измеряемых параметрах. Наличие постоянно обновляемых оперативных данных особо ценно при решении задач ассимиляции и задач прогностических расчетов в акваториях Мирового океана. Все виды оперативных данных, представленных на портале проекта, обновляются ежедневно, что позволяет использовать данные для прогностических расчетов в реальном времени при наличии соответствующей численной модели (Agoshkov et al., 2015). Если же говорить о частоте данных по времени, так называемом временном разрешении, то о температуре поверхности Балтийского моря на портале проекта представлены среднесуточные данные и данные за каждые три часа на регулярных сетках или данные наблюдений in-situ. Как правило, для корректировки численной модели достаточно усвоения среднесуточных данных. Для проведения прогностических расчетов – данных о состоянии ассимилируемых параметров каждые три часа. Треки чаще используются для локальных моделей и проверки некоторых из параметров численных моделей моря. Частота представленных данных достаточна для ассимиляции данных наблюдений в численной модели термодинамики Балтийского моря, для проведения корректировки модели или расчета краткосрочных и среднесрочных прогнозов.

Также стоит отметить, что данные, представленные на сайте проекта, имеют достаточно высокое пространственное разрешение. Разрешение данных моносенсоров и мультисенсорных данных наблюдений (Sea surface temperature Multi-sensor L3 observations) составляет $0,02 \times 0,02$ градуса. В численной модели термодинамики Балтийского моря, разрабатываемой в ИВМ РАН (Zalesny et al., 2014), пространственное разрешение составляет $0,0625 \times 0,03125$ градуса. Что означает, что на данный момент пространственное разрешение представленных на указанном ресурсе данных наблюдений высоко и достаточно для проведения численных экспериментов по ассимиляции данных в численной модели термодинамики Балтийского моря без внесения дополнительных ошибок (которые могут появляться при недостаточном количестве данных).

Для устранения ошибок и погрешностей измерений данные, прежде чем выложить на ресурс, калибруют на основе параметров измерительных приборов и валидируют в центрах обработки информации. Но даже после проведения автоматических процедур калибровки и валидации данных в них могут оставаться ошибки, которые в дальнейшем при процедурах ассимиляции данных могут повлиять на численное решение математической модели исследуемой акватории (Gejadze, Le Dimet, Shutyaev, 2011; Shutyaev et al., 2012). Некоторые ошибки могут быть замечены только при детальном рассмотрении получаемых данных и анализе данных, основанном на физических свойствах исследуемой акватории (Гандин, Каган, 1976). Дополнительная верификация данных, описанная в настоящей работе, проводилась на основе данных наблюдений о температуре поверхности Балтийского моря. В данных «мультисенсоров», получаемых с портала проекта Коперникус при проведении мониторинга ТПМ в акватории Балтийского моря, были замечены значения, правдоподобность которых подверглась сомнению. На *рис. 1* представлены данные наблюдений «мультисенсоров» о температуре поверхности Балтийского моря (Copernicus Product ID SST_eur_sst_

L3S_nrt_observations_010_009_a) на разные моменты времени 2 февраля 2015 года. Из сравнения четырех последовательных полей ТПМ видно, что значения температуры на 6 часов утра указанной даты сильно отклоняются от соседних по времени значений температур, и в Финском заливе ТПМ превышает 14°C там, где в остальные моменты времени не достигает и 6°C . Возможность такого краткосрочного нагрева (и затем охлаждения) части акватории крайне маловероятна, поэтому анализу и проверке указанных данных наблюдений было уделено повышенное внимание.

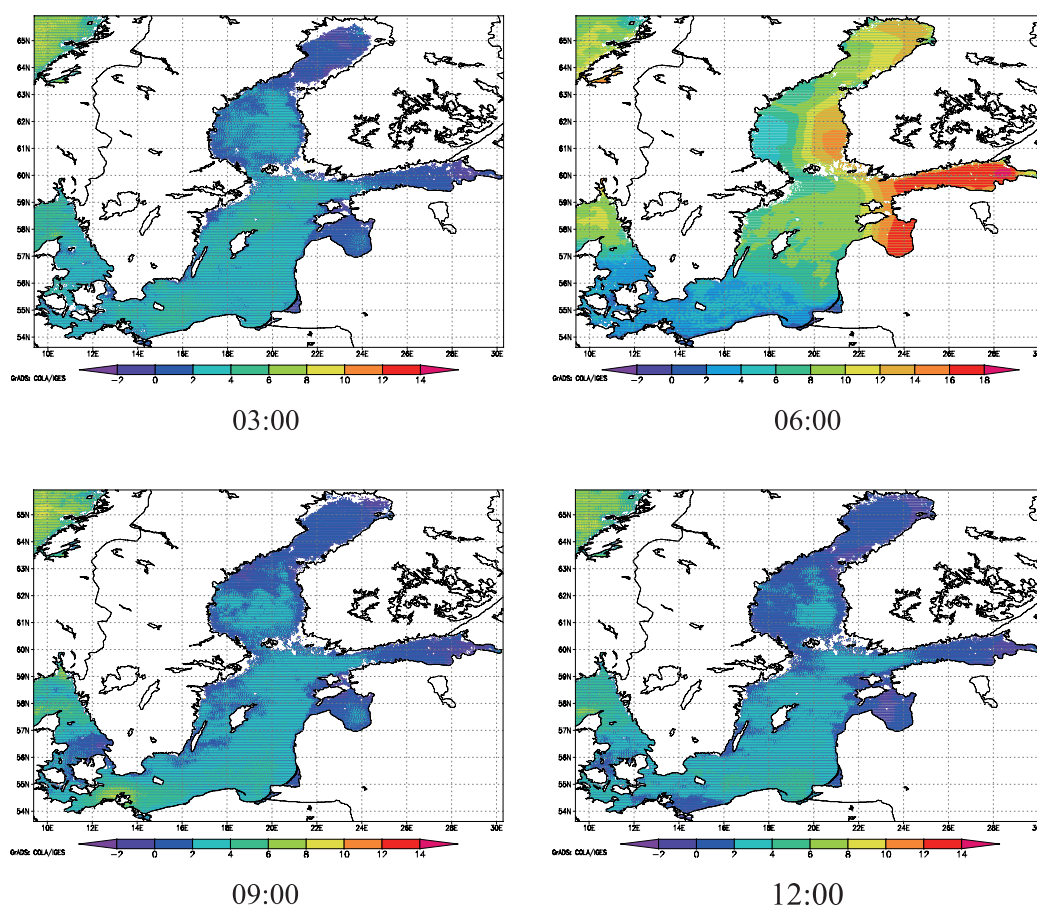


Рис. 1. Температура поверхности Балтийского моря 2 февраля 2015 г. по данным наблюдений на разные моменты времени

Для дополнительного анализа поступивших данных было проведено сравнение с данными Гидрометцентра России (<http://meteoinfo.ru/>) о среднесуточной ТПМ Балтийского моря на 18 февраля 2015 года. Исходя из анализа среднесуточных данных и учитывая, что даже в самые теплые месяцы в году поверхность Балтийского моря не прогревается за день на $8\text{--}10^{\circ}\text{C}$ (Karagali, Noyer, Nasager, 2012), был сделан вывод о том, что исследуемые данные содержат ошибку. Более того, анализ данных выявил регулярные повторения данной ошибки, что говорит о ее возможной приборной составляющей.

Учитывая вышесказанные замечания, приходим к выводу, что даже после калибровки и валидации данных наблюдений в вычислительных центрах обработки информации до выставления на сайт для пользователей данные могут содержать ошибку. Поэтому, помимо процедур интерполяции данных на расчетные сетки численных моделей (Zakharova,

Agoshkov, Parmuzin, 2013), появляется необходимость проводить дополнительную проверку данных наблюдений на «физичность». Поскольку делать такую верификацию вручную постоянно не представляется возможным, требуется создание автоматических процедур дополнительной проверки данных наблюдений.

Верификация данных

Для дополнительной верификации данных наблюдений о ТПМ был опробован статистический подход. Это позволяет вместо исследования индивидуальных свойств отдельно взятых полей путем рассмотрения их статистических характеристик установить общие особенности, характерные для всего набора реализаций полей температур. Эти общие особенности принято называть статистической структурой случайного поля.

Вводится статистическое осреднение \bar{f} , которое характеризует среднее из возможных значений в данной точке в данный момент времени и определяется по формуле:

$$\bar{f}(r) = \frac{1}{N} \sum_{j=1}^N f^{(j)}(r),$$

где $f^{(j)}$ – j -тая реализация поля ТПМ, r – радиус-вектор точки (причем среди координат могут быть как пространственные координаты, так и время), N – число реализаций.

Теперь наибольший интерес представляет рассмотрение величин, описывающих отклонение f от среднего. Для оценки возможного разброса вводится величина, представляющая собой средний квадрат отклонения f от средней величины:

$$\sigma_f^2(r) = \frac{1}{N} \sum_{i=1}^N (f_i(r) - \bar{f}(r))^2.$$

Величина σ_f называется средним квадратическим отклонением (среднеквадратичным отклонением) величины f .

Далее, для верификации данных наблюдений (и отброса «нефизичных» полей) можно ввести критерий на основе известного в теории вероятностей и статистике Правила трех сигм: «Если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратического отклонения» (Гмурман, 2003). Проще говоря и переходя к терминам нашей задачи, в соответствии с правилом, вероятность того, что абсолютная величина отклонения от среднего статистического превысит утроенное среднее квадратическое отклонение, очень мала.

Стоит отметить, что поля температуры поверхности моря, вообще говоря, нельзя считать нормально распределенной случайной величиной, но правило трех сигм может служить основой к методу дополнительной верификации данных и выполняется на большей части акватории. Для подтверждения этого был проведен ряд численных экспериментов.

Результаты численных расчетов

В настоящей работе для проведения численных расчетов использовались данные сервиса Коперникус по мониторингу и моделированию морских сред (<http://marine.copernicus.eu/>). По данным о температуре поверхности Балтийского моря за 25–27 лет были посчитаны статистическое осреднение и среднее квадратическое отклонение, на основе которых проводилась верификация данных наблюдений со спутников о ТПМ за 2015 год согласно описанному выше подходу. Статистическое осреднение \bar{T} поля температуры поверхности Балтийского моря было посчитано по данным реанализа о среднемесячной ТПМ с использованием спутниковой информации с инфракрасных радиометров за 1989–2013 года. (Copernicus Product ID BalticSea_reanalysis_phys_003_008). На *рис. 2* приведено вычисленное среднее статистическое поле ТПМ за февраль, от которого в дальнейшем высчитывалось отклонение значений полей температуры поверхности, «физичность» которых ставилось целью определить.

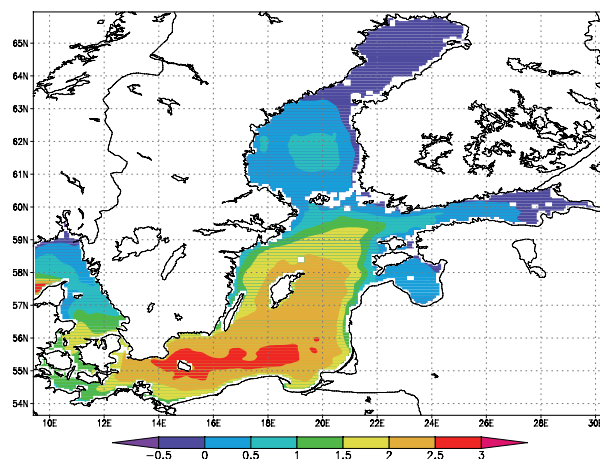


Рис. 2. Статистическое осреднение ТПМ по данным за 1989–2013 гг., февраль

Далее, по данным реанализа о среднесуточной ТПМ с 1982 г. по 2009 г. (Copernicus Product ID SST_Bal_SST_L4_rep_observation_010_016) было вычислено значение среднего квадратического отклонения σ (см. *рис. 3*).

Теперь, согласно правилу трех сигм, значения ТПМ на большей части акватории должны удовлетворять следующему критерию:

$$|T_i - \bar{T}| \leq 3\sigma.$$

Можно переписать этот критерий в следующем виде:

$$D_i \equiv 3\sigma - |T_i - \bar{T}| \geq 0,$$

т.е. поле D_i должно быть неотрицательно на большей части исследуемой акватории. Введение такого обозначения позволяет более наглядно изобразить результаты численных расчетов по оценке поля ТПМ и его проверки на «физичность». На *рис. 4* приведены

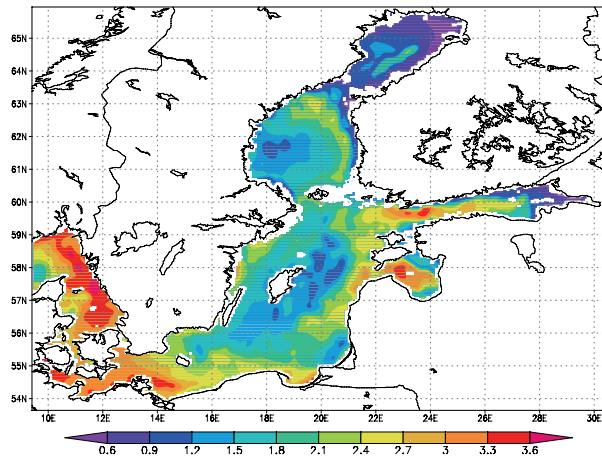


Рис. 3. Среднеквадратическое отклонение ТПМ Балтийского моря

поля D_i , вычисленные на основе оцениваемого поля ТПМ на 18 февраля 2015 года в различные моменты времени. Серым цветом на рисунке помечены положительные значения поля D_i , черным – отрицательные.

Из сопоставления рис. 1 и рис. 4 видно, что поле, не вызывающее изначально сомнения, имеет по большей части положительное поле D_i , что подтверждает выкладки, сделанные на основе правила трех сигм. В то же время поле ТПМ, которое не проходит проверку по внешним признакам, имеет в большинстве своем отрицательные значения поля D_i , что говорит о том, что данное поле наблюдений «не физично». Такое поле следует воспринимать ошибочным, а следовательно, не ассимилировать его в численной модели во избежание внесения ошибок. Таким образом, проводим верификацию данных наблюдений и отбрасываем поля, не прошедшие проверку, перед тем, как направлять данные в численную модель для ассимиляции и проведения дальнейших расчетов.

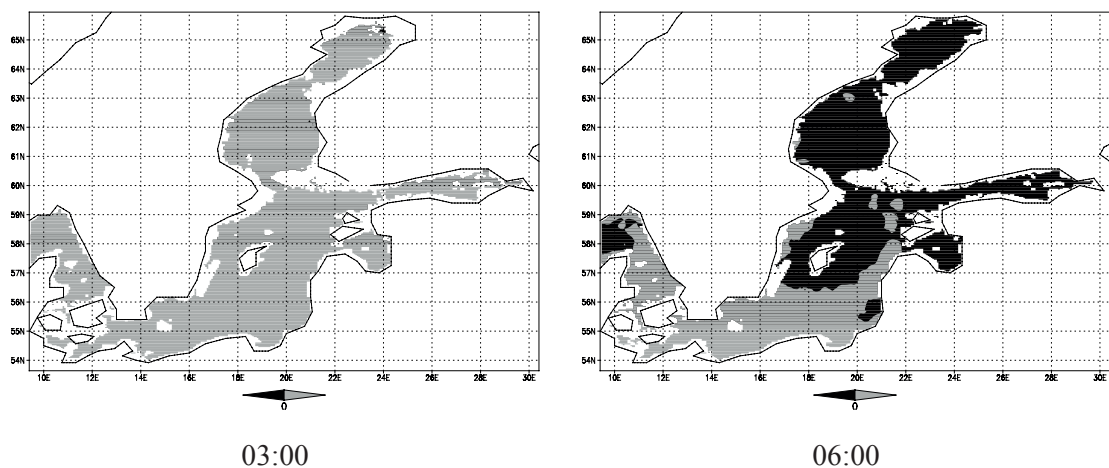


Рис. 4. Поле D , вычисленное для данных наблюдений ТПМ Балтийского моря на 18 февраля 2015 года в моменты времени 6:00 и 9:00 часов

Заключение

В настоящей работе для дополнительной верификации данных наблюдений о ТПМ реализован статистический подход, основанный на правиле трех сигм. Представлены результаты его применения на примере данных со спутников о температуре поверхности Балтийского моря. Метод подходит для дополнительной верификации данных и исключения, таким образом, нефизичных полей температур.

Дополнительной сложностью применения такого метода является необходимость при вычислении статистических значений, задействованных в предлагаемом методе, постоянно дополнять базу данных актуальными значениями. Это означает, что для верификации оперативных данных следует не только посчитать, с использованием статистики, требуемые осреднения, участвующие в критерии отбора полей температур (что было сделано в настоящей работе), но и вести базу данных для расчета актуальных осреднений.

Также остается вопрос, как не пропустить при таком подходе действительно происходящие сильные изменения исследуемой среды. Этот вопрос актуален в случаях чрезвычайных ситуаций и аномальных скачков температур.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 15-01-01583, в рамках которого проводилось исследование ошибок данных наблюдений) и Российского научного фонда (проект 14-11-00609, в рамках которого проводились исследовательские испытания комплекса программ по пересчету данных наблюдений о температуре поверхности Балтийского моря).

Литература

1. Гандин Л.С., Каган П.Л. Статистические методы интерпретации метеорологических данных. Л.: Гидрометеоздат, 1976. 359 с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высш. шк., 2003. 479 с.
3. Agoshkov V.I., Parmuzin E.I., Zakharova N.B., Zalesny V.B., Shutyaev V.P., Gusev A.V. Variational assimilation of observation data in the mathematical model of the Baltic Sea dynamics // Russian Journal of Numerical Analysis and Mathematical Modelling. 2015. Vol. 30. Issue 4. P. 203–212.
4. Gejadze I., Le Dimet F.-X., Shutyaev V. Computation of the optimal solution error covariance in variational data assimilation problems with nonlinear dynamics // Journal of Computational Physics. 2011. Vol. 230. P. 7923–7943.
5. Karagali I., Hoyer J., Hasager C.B. SST diurnal variability in the North Sea and the Baltic Sea // Remote Sensing of Environment. 2012. V. 121. P. 159–170.
6. Shutyaev V.P., Le Dimet F.-X., Gejadze I.Yu., Copeland G.J.M. Optimal solution error covariance in highly nonlinear problems of variational data assimilation // Nonlinear Processes in Geophysics. 2012. Vol. 19. P. 177–184.
7. Zakharova N.B., Agoshkov V.I., Parmuzin E.I. The new method of ARGO buoys system observation data interpolation // Russian Journal of Numerical Analysis and Mathematical Modelling. 2013. Vol. 28. Issue 1. P. 67–84.
8. Zalesny V.B., Gusev A.V., Chernobay S.Yu., Aps R., Tamsalu R., Kujala P., Rytkönen J. The Baltic Sea circulation modelling and assessment of marine pollution // Russian Journal of Numerical Analysis and Mathematical Modelling. 2014. Vol. 29. Issue 2. P. 129–138.

Verification of the sea surface temperature observation data

N.B. Zakharova

Institute of Numerical Mathematics RAS, Moscow 119333, Russia

E-mails: zakharova_nb@inm.ras.ru

The paper is devoted to processing of hydrophysical observations. Unfortunately, there may be errors in data even after calibration and validation made by information centers. Therefore, it is necessary to conduct an additional verification of the observations before using them in a range of hydrodynamics problems. In this work, the verification is conducted by an example of satellite sea surface temperature data (SST). An analysis of the Baltic SST observations was made with the identification of errors in the data on the basis of physical properties of the water area.

The method of additional verification of the satellite data is described. It is based on statistical approaches and allows one to set general features that are typical of the entire set of temperature field realizations. The results of this method applied to the real-time data are presented. Numerical experiments are conducted using statistical data on SST. Daily and monthly reanalysis data from the Copernicus marine environment monitoring service have been used in the calculations. The implemented method helps to solve the problem of the data verification by means of excluding the incorrect data fields.

Keywords: mathematical modeling, satellite observation data, data verification, sea surface temperature

Accepted: 10.05.2016

DOI: 10.21046/2070-7401-2016-13-3-106-113

References

1. Gandin L.S., Kagan R.L., *Statisticheskie metody interpretatsii meteorologicheskikh dannykh* (Statistic methods of the meteorological data interpretation), St. Petersburg: Hydrometeoizdat, 1976, 359 p.
2. Gmurman V.E., *Teoriya veroiatnostey i matematicheskaya statistika* (Probability theory and mathematical statistics). Moscow: Vysshaya shkola, 2003, 479 p.
3. Agoshkov V.I., Parmuzin E.I., Zakharova N.B., Zalesny V.B., Shutyaev V.P., Gusev A.V., Variational assimilation of observation data in the mathematical model of the Baltic Sea dynamics, *Russian Journal of Numerical Analysis and Mathematical Modelling*, 2015, Vol. 30, Issue 4, pp. 203–212.
4. Gejadze I., Le Dimet F.-X., Shutyaev V., Computation of the optimal solution error covariance in variational data assimilation problems with nonlinear dynamics, *Journal of Computational Physics*, 2011, Vol.230, pp. 7923–7943.
5. Karagali I., Hoyer J., Hasager C.B., SST diurnal variability in the North Sea and the Baltic Sea, *Remote Sensing of Environment*, 2012, Vol. 121, pp. 159–170.
6. Shutyaev, V.P., Le Dimet, F.-X., Gejadze, I.Yu., Copeland, G.J.M., Optimal solution error covariance in highly nonlinear problems of variational data assimilation, *Nonlinear Processes in Geophysics*, 2012, Vol. 19, pp. 177–184.
7. Zakharova N.B., Agoshkov V.I., Parmuzin E.I., The new method of ARGO buoys system observation data interpolation, *Russian Journal of Numerical Analysis and Mathematical Modelling*, 2013, Vol. 28, No. 1, pp. 67–84.
8. Zalesny V.B., Gusev A.V., Chernobay S.Yu., Aps R., Tamsalu R., Kujala P., Rytönen J., The Baltic Sea circulation modelling and assessment of marine pollution, *Russian Journal of Numerical Analysis and Mathematical Modelling*, 2014, Vol. 29, Issue 2, pp. 129–138.