

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Development of a Chemical Source Apportionment Decision Support Framework for Catchment Management

Journal:	<i>Environmental Science & Technology</i>
Manuscript ID:	es-2013-01793e.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Comber, Sean; Plymouth University, Environmetnal Science Gardner, Michael; Atkins Limited, Daldorph, Peter; Atkins Limited, Constantino, Carlos; Atkins Limited, Smith, Russell; Westcountry Rivers Trust, Ellor, Brian; UKWIR,

SCHOLARONE™
Manuscripts

1 **Development of a Chemical Source Apportionment Decision Support**
2 **Framework for Catchment Management**

3
4
5 Sean D W Comber*, Russell Smith,¹ Peter Daldorph², Michael J Gardner², Carlos
6 Constantino², Brian Ellor³

7
8 * corresponding author: B525 Portland Square, Plymouth University, Drake Circus,
9 Plymouth, PL4 8AA, UK. Tel: +44(0)1752858974. Fax: +44(0)1752584710. Email:
10 sean.comber@plymouth.ac.uk

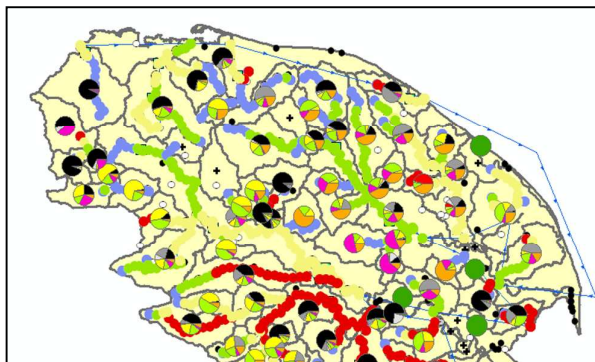
11 ¹ Westcountry Rivers Trust, Rain-Charn House, Stoke Climsland, Cornwall, PL17
12 8PH.

13 ² Atkins Limited, The Hub, 500 Park Avenue, Aztec West, Bristol, BS32 4RZ,
14 UK.

15 ³ UK Water Industry Research, 1 Queen Anne's Gate, London SW1H 9BT, UK.
16
17

18 **Research Article – Environmental Modeling**

19
20
21 **Abstract Art**
22



23
24
25
26 **Key words**

27
28 Chemical, source apportionment, nutrients, metals, environmental modeling
29
30
31
32
33

34 **Abstract**
35

36 EU legislation, including the Water Framework Directive, has led to the application of
37 increasingly stringent quality standards for a wide range of chemical contaminants in
38 surface waters. This has raised the question of how to determine and to quantify the
39 sources of such substances so that measures can be taken to address breaches of
40 these quality standards using the polluter pays principle. Contaminants enter surface
41 waters via a number of diffuse and point sources. Decision support tools are required
42 to assess the relative magnitudes of these sources and to estimate the impacts of
43 any programmes of measures. This paper describes the development and testing of
44 a modeling framework, the Source Apportionment Geographical Information System

45 (SAGIS). The model uses readily available national data sets to estimate
46 contributions of a number of nutrients (nitrogen and phosphorus), metals (copper,
47 zinc, cadmium, lead, mercury and nickel) and organic chemicals (a phthalate and a
48 number of polynuclear aromatic hydrocarbons) from multiple sector sources. Such a
49 tool has not been available on a national scale previously for such a wide range of
50 chemicals. It is intended to provide a common platform to assist stakeholders in
51 future catchment management.

52

53

54 Introduction

55

56 The European Union Water Framework Directive (WFD)¹ sets criteria for water
57 bodies to meet a defined status categorised as 'good', which requires chemical
58 standards for 33 priority and priority hazardous substances and groups of substances
59 to be achieved, in addition to standards for ecology, hydrology and hydromorphology.
60 The Environmental Quality Standards (EQSs) set for these substances are generally
61 more stringent than existing EQSs that were under other previous Directives,
62 including the Dangerous Substances Directive.²

63

64 Historically, the principal method for improving river water quality has been to place
65 restrictive discharge consents on point source discharges, including those from
66 wastewater treatment works. In the UK, a combination of the development of a less
67 polluting industrial base as a response to more stringent regulation in the form of EU
68 Directives and downward pressure on limit values from other point sources, means
69 that for a number of substances, diffuse sources from agriculture, urban runoff, soil
70 erosion and discharges from abandoned mines make an increasingly important
71 contribution to exceedances of EQSs.³ The WFD advocates the application of the
72 'polluter pays principle'¹ to ensure that any one sector is not unduly burdened with
73 the requirement to reduce discharges to meet an EQS. To plan to meet the stringent
74 standards set by the Directive it is therefore necessary to quantify the significance of
75 all sources to an EQS exceedance in any given water catchment. In many instances
76 it is likely that mitigation measures targeting multiple sources / sectors will be needed
77 to achieve good ecological status, so a clear appreciation of relative contributions is
78 essential.

79

80 In order to establish plans to comply with new EQSs it is necessary for regulators
81 and regulated alike to have tools to test and support planning decisions. A number of
82 models have been developed to predict inputs of chemicals from agricultural diffuse
83 sources; notable amongst these in the UK are PSYCHIC for soil and phosphorus and
84 NEAP-N for nitrogen. NEAP-N is a simple model created by ADAS that looks at
85 leachate from different land uses.⁴ It incorporates details down to for example a
86 livestock or fertilizer management level. Output is a visual representation of grid
87 squares over the catchment of interest, representing predicted change in
88 concentrations per hectare. The model does not include a within stream fate
89 component, so it is primarily used as a source of information on agricultural inputs of
90 nitrogen into a catchment based water quality model. PSYCHIC, The Phosphorus

91 and Sediment Yield CHaracterisation In Catchments model has been developed by
92 ADAS, NSRI, CEH-Wallingford, and the Universities of Exeter, Reading and Sheffield
93 through a Defra funded research programme.⁵ PSYCHIC was developed to model
94 phosphorus (P) and sediment transfers in agricultural systems to investigate options
95 to reduce P delivery from land to rivers, particularly through identification of hotspots
96 and an associated field-scale risk assessment exercise.

97
98 Once discharged to a watercourse, any given chemical will be subject to dilution and
99 undergo various biogeochemical processes that might be incorporated into a model.
100 Water Quality models include the United States Environmental Protection Agency
101 (USEPA) model QUAL2E;⁶ the MIKE series of models that is developed by the
102 Danish Hydraulics Institute, and the Systeme Hydrologique European (SHE)⁷; The
103 most used models by the Environment Agency of England and Wales are SIMCAT
104 and TOMCAT. SIMCAT is able to simulate a statistical distribution of discharge and
105 water quality data for multiple effluent inputs within a catchment. It is capable of
106 simulating up to 2500 random boundary conditions (also known as the Monte Carlo
107 approach), based on the input data, SIMCAT produces a distribution of results from
108 which an assessment of the impact can be made on the predicted mean and ninety-
109 five percentile concentrations.⁹ SIMCAT allows for inputting decay constants based
110 on a first order decay rate and provides options for point and diffuse source inputs. In
111 addition the SIMCAT source code is sufficiently flexible to allow upgrades to include
112 such options as inputting partition coefficients to allow concentrations of, for example,
113 metals, to be split into dissolved and particulate bound concentrations.

114
115 Regulators have also sought to develop screening models for assessing pollutant
116 pressure in order to plan relevant measures on a national scale. The Scottish
117 Environmental Protection Agency (SEPA) have developed such a screening tool; the
118 Diffuse Pollution Screening Tool (DPST)¹⁰ which has drawn together large national
119 datasets for a number of different types of chemicals including metals, nutrients,
120 pesticides and sanitary determinands (BOD, ammonia). However, the focus of the
121 model is on source apportionment alone, rather than predicting in-river
122 concentrations, partly owing to the fact that there is no national water quality model
123 for Scotland, akin to that of SIMCAT for England and Wales.

124
125 The difference between source apportionment and water quality prediction is a key
126 distinction. Source apportionment models have value in risk assessment of
127 determining input loads and to some degree locations, but their application to water

128 quality modeling is not straightforward. Modeling of the mixing of inputs, taking
129 account of the variability of flow and load and incorporating appropriate processes of
130 chemical behaviour add a considerable level of complexity. There are currently few
131 models which can do this on a local or regional scale and none on a national scale.
132 One model developed to achieve catchment management of a regional scale is the
133 GREAT-ER model (Geographically-referenced Regional Exposure Assessment Tool
134 for European Rivers). The model was originally developed to assess the exposure
135 risk of new substances discharged to sewer from predominantly domestic sources,
136 but has been expanded to include diffuse and point sources within a number of
137 German catchments. GREAT-ER is a hybrid Monte-Carlo deterministic model which
138 allows a user to calculate the predicted environmental concentration (PEC) for a
139 substance, taking into account its geographical and temporal distribution, to produce
140 a statistical output and perform 'what-if' scenarios Data entry and selection is via a
141 GIS interface and In the receiving watercourse, chemical processes are represented
142 by conservative dilution, first order decay (similar to SIMCAT) or more complex
143 processes (similar to QUAL2E). Output is provided as annual statistics produced
144 longitudinally downstream and converted into PEC values, which are displayed via
145 the GIS interface. The modeling system is open-source ([http://www.great-
146 er.org/pages/home.cfm](http://www.great-er.org/pages/home.cfm)) with the aim of creating a live development framework. The
147 GREAT-ER model has been recently applied to the Ruhr catchment for the source
148 apportionment of zinc which includes background inputs, discharges from mining
149 activities, runoff, sewage treatment plant sources and diffuse agricultural
150 contributions.¹¹

151

152 The development of river basin management plans to meet WFD objectives requires
153 the assessment of a synthesis of local and national, point and diffuse source
154 measures, this paper describes a modeling framework developed to utilise for the
155 first time, national datasets for multiple parameters including hydrology, rainfall,
156 modelled discharges of chemicals, reported discharge loads, and spatial datasets
157 including the locations of wastewater treatment works and smaller on-site works
158 often septic tanks, combined sewer overflow locations, output from diffuse pollution
159 risk models, road and river system networks combined within a GIS based modeling
160 framework to provide estimates of pollutant loads and in-river concentrations of
161 chemicals at water body scale for the whole of England and Wales.

162

163

164

165 **Methods**

166

167 The Source Apportionment-GIS (SAGIS) modeling framework integrates information
168 from multiple sources. Loads from different source sectors are derived from
169 coefficients expressed as an annual or monthly mean load with corresponding
170 standard deviation. Point sources are represented as mean and standard deviations
171 of concentrations and flow with the option of breaking down inputs into monthly
172 values to allow simulation of seasonal effects. Diffuse sources are represented as
173 mass per year, or month, per km². All loads are routed into associated river reaches
174 using one of 18 regional SIMCAT models covering England and Wales. Simcat
175 models are being developed for Scotland and so currently only loads can be derived.
176 To cover the whole of Great Britain (England, Wales and Scotland) on this basis, the
177 use of datasets with national coverage was imperative to provide both consistency of
178 approach and the ability to manage and update data. A common map projection was
179 used for all databases and mapping based on a 1km² grid of England, Wales and
180 Scotland. Such GIS mapping calibration and validation had previously been
181 undertaken as part of previous projects associated with the hydrological and diffuse
182 source components. Detailed information regarding the methodologies used to
183 calculate loads for each source is provided elsewhere.¹² However, a brief description
184 of the data and method used to derived load estimates is provided in the following
185 section.

186

187 Table 1 summarises the type of source (both diffuse and point) and notes whether
188 loads to surface waterbodies were either derived via an established model or were
189 calculated as part of this research. Figure 1 provides a schematic for the structure
190 and key components of the SAGIS decision support framework.

191

192 **Table 1** **Summary of methodology used to calculate loads discharged to**
 193 **waterbodies**
 194

Category	Source	Metals	Nutrients	Organics
Agriculture – Arable and Livestock	Diffuse	n/a ¹	PSYCHIC (P), NEAP-N models	n/a
Highway (non urban runoff)	Diffuse	Highway Agency HAWRAT model	Highway Agency HAWRAT model	Highway Agency HAWRAT model
Urban runoff	Diffuse	Calculated	Calculated	Calculated
Background erosion	Diffuse	PSYCHIC model + calculated	n/a	PSYCHIC model + calculated
Onsite wastewater treatment systems	Diffuse	Environment Agency model ² + calculated	Calculated	Calculated
Atmospheric deposition	Diffuse	Calculated	P n/a N within NEAP-N models	Calculated
Treated wastewater effluent	Point	Measured ³ and defaults	Measured and defaults	Measured and defaults
Storm tanks/combined sewer overflows	Point	Calculated	Calculated	Calculated
Industrial discharges	Point	Environment Agency Measured/reported	Measured/reported	Measured/reported
Mine water discharges	Point	Environment Agency Measured/reported	n/a	n/a

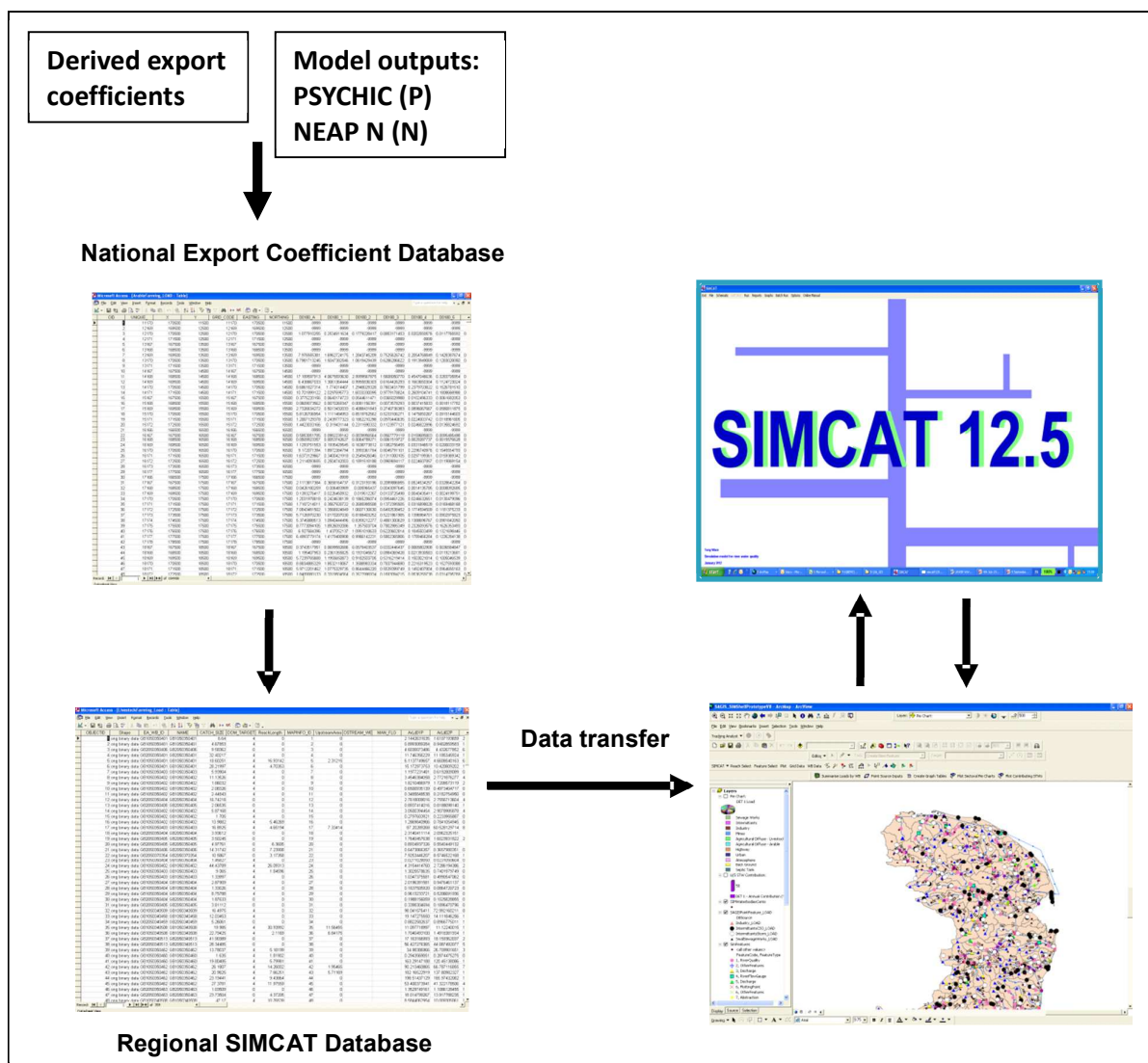
195 ¹ Included in erosion category; ² EA model used to predict locations of onsite wastewater
 196 treatment systems

197 ³ based on EA and Water Company data (flow and concentration);

198

199

200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215



216 **Figure 1 Schematic diagram for SAGIS tool structure**

217

218 Tables which list the key datasets used to derive the exported loads from each
219 source and information as to how the datasets were used to derive the calculated
220 loads to waterbodies are provided in Table 1 and 2 of the Supporting Information
221 respectively.

222 Owing to WwTW inputs being significant for many substances of interest, it was
223 critical to utilise as much monitoring data as possible to derive accurate loads
224 entering receiving waters. In the absence of effluent data for WwTW, ie for WwTW

225 where there is not a requirement to determine the chemicals of interest in their
226 effluent owing to the absence of a consent, then default values were required. All
227 available concentration data were collected and collated for Environment Agency
228 monitoring between 2007 and 2009 inclusive, which added up to a maximum of over
229 2,000 results for phosphorus and nitrogen, several hundred for the metals to very few
230 for the organic determinands. In most cases a mean concentration was chosen as
231 the default value. For phosphorus because the non-consented works without effluent
232 data would not have had phosphorus reduction measures installed, a default value
233 was generated from all monitoring data reported above 2 mg-P/l. Details of the
234 monitoring data used and the default values chosen are provided in Table 3 of the
235 Supporting Information.

236 Inputs to SIMCAT for point sources also require a flow for WwTW effluent discharges
237 in order to generate a load.

238

239 Flows for the works were based on a number of collated data in the following
240 hierarchy of available data:

- 241 1) Measured flows and standard deviations provided by the water companies
- 242 2) Consented DWF
- 243 3) Populations multiplied by an average flow per capita per day (from all
244 sources) assumed to be 250 l/capita/day.

245 A summary of the number of WwTW applicable to each category is provided in Table
246 4 of the Supporting Information.

247 Reported literature runoff data¹⁵ was used to generate concentrations of the
248 substances of interest in road runoff expressed as an event mean concentration
249 (which takes account of the 'first flush' effect and subsequent reduction in
250 concentration with increasing rainfall) and in domestic dry weather flow (Table 2).

251

252 **Table 2** Runoff and dry weather flow concentration data¹³

Concentrations used	Event Mean Concentration for Road Runoff (µg/l)	Domestic raw sewage (µg/l)
Copper	34.7	186.1
Zinc	82.5	62.6
Total PAHs	0.68	0.76
Fluoranthene	0.06	0.093
Naphthalene	0.08	0.077
Benzo(a)pyrene	0.03	0.025
Benzo(b)fluoranthene	0.03	0.030
Benzo(k)fluoranthene	0.04	0.030
Benzo(ghi)pyrene	0.03	0.025
Indeno(123-cd)pyrene	0.03	0.025
DEHP	20	33.1
Total Nitrogen (mg/l)	1.93	39.7
Total Phosphorus (mg/l)	0.27	13.9

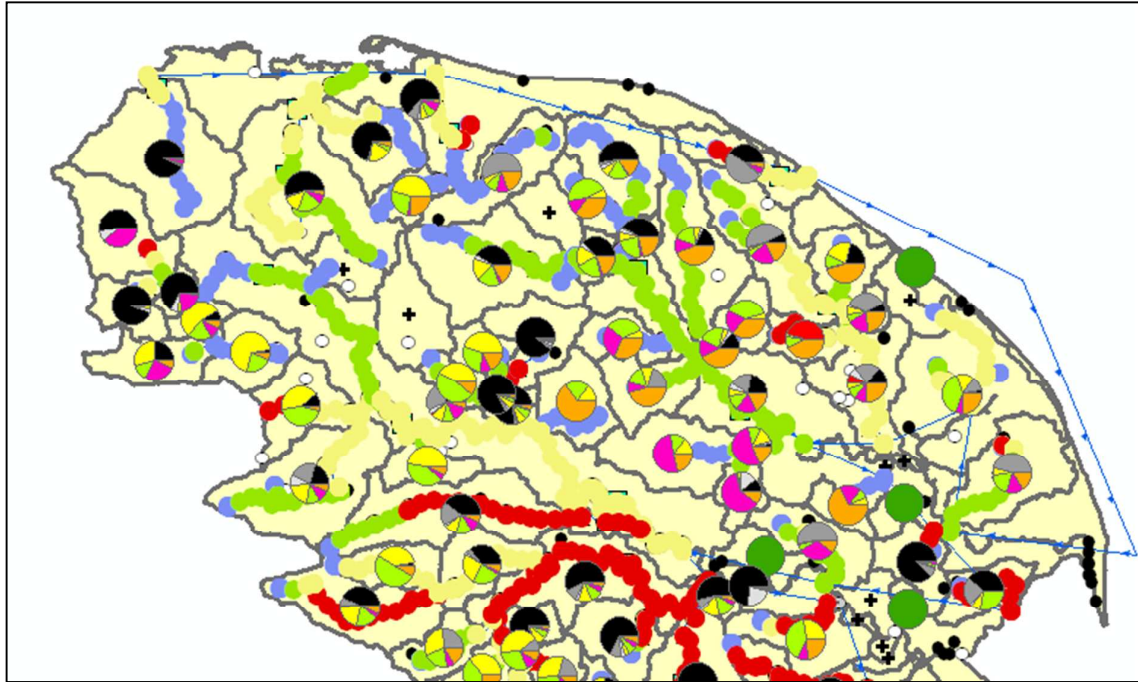
253

254 Water quality monitoring data provided by the Environment Agency and SEPA was
255 used for model testing purposes.

256

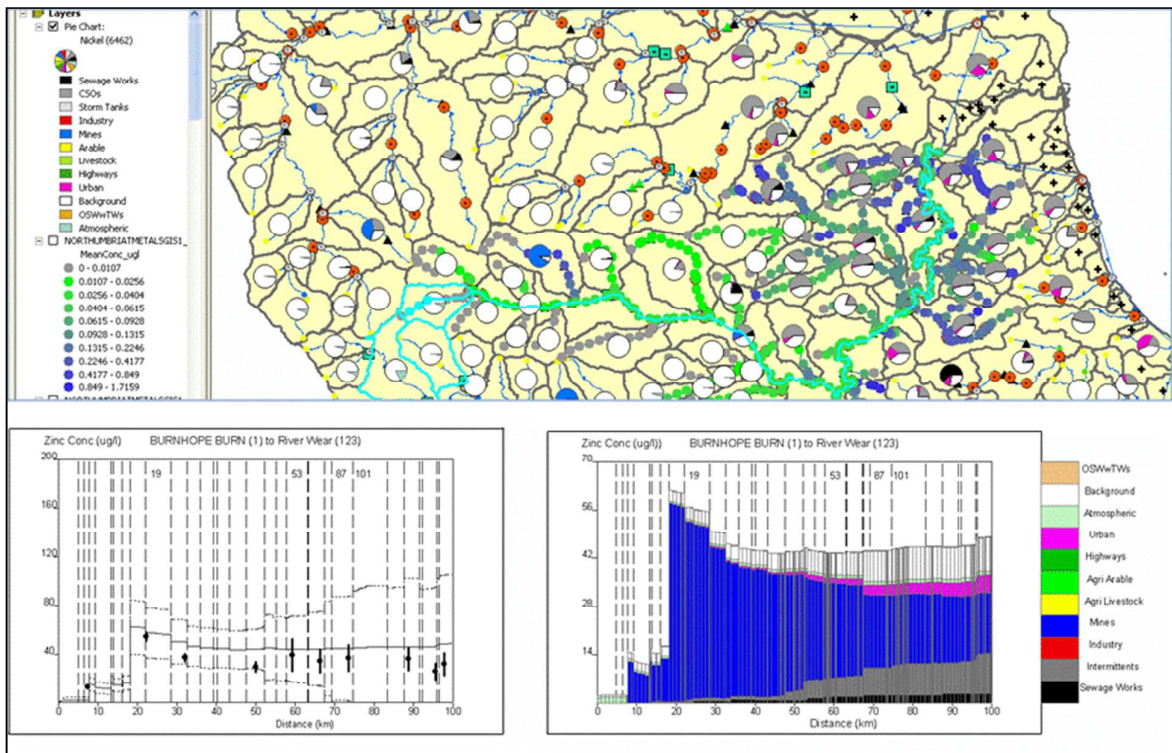
257 Each source was represented within a Microsoft Access™ database either as a point
258 source with an X and Y, UK national grid location coordinate or as an individual 1 km²
259 grid (approximately 150,000 for England and Wales) (Figure 1). The main database
260 is then split into 18 regional Access databases (see Figure 1 in the Supporting
261 Information) which form the attribute tables behind the features in ArcMAP 9.3™ GIS
262 software. The functionality with ArcMap and bespoke macros developed in Visual
263 Basic are then used to extract the necessary data and generate the text file required
264 to run SIMCAT, a stochastic water quality model. SIMCAT can be run from within
265 SAGIS and provide outputs (total and dissolved concentrations and loads for metals,
266 total concentrations and loads for nutrients and organics) which are fed back into
267 ArcGIS to provide cartographic, graphical and spreadsheet outputs (Figures 2 and 3).

268



269
270
271
272
273

Figure 2 Example source apportionment from SAGIS for phosphorus in East Anglia



274
275
276

Figure 3 Example outputs from SAGIS

277 **Results**

278 The model was initially populated with input data for total nitrogen, total phosphorus,
279 total copper, zinc, nickel, lead, cadmium and lead, PAHs (naphthalaene, anthracene,
280 fluoranthene, benzo-a-pyrene, benzo-b-fluoranthene, benzo-k-fluoranthene, benzo-
281 ghi-perylene, indeno-123,cd-pyrene) and diethylhexylphthalate. Extension to other
282 substances is under ongoing consideration.

283

284 The SAGIS model provides a number of outputs (Figures 2 and 3), including:

- 285 • Colour coded concentrations within the river system at 1km intervals which
286 can be aligned with compliance assessment guidelines (e.g. EQS or
287 ecological status under the WFD)
- 288 • Pie charts illustrating the relative contributions from all upstream sources to
289 the load or concentration at the outflow from each waterbody
290 (approximately 7,000 in England and Wales)
- 291 • Pie charts illustrating the relative contributions from different sources to the
292 load on a waterbody basis
- 293 • Cumulative concentration from each source along a river length defined by
294 the user
- 295 • Predicted versus observed concentrations where monitoring data is
296 available

297

298 The ability to graphically present the percentage contribution from the different point
299 and diffuse source sectors provides a very visual representation of the main
300 contributors to loads or concentrations of a chemical to any given waterbody. Such
301 outputs are vital in engaging stakeholders in the process of improving water quality
302 under the Water Framework Directive.

303

304 When integrating models from many sources for different components of physic-
305 chemical environments, with varying spatial and temporal specificity, the question of
306 parameter estimation (i.e. accuracy of export coefficients) and potential error
307 propagation becomes paramount. To test the predictive skill of SAGIS based on the
308 National / default data layers a comparison between predicted and measured
309 concentration data was undertaken for three selected catchments of differing
310 typology (mine dominated – river Wear in the NE of England, part of the
311 Northumbrian model and mixed, urban – river Tame Midlands, part of the river Trent
312 model and rural river Avon part of the SW England model) (see Figure 2 in the
313 Supporting Information).

314

315 The process aimed to:

- 316 • Identify any data transfer errors in the development of the export coefficient
317 databases and the input of these to SIMCAT
- 318 • Identify where the default export coefficients based on national datasets do
319 not provide a good representation in individual catchments and may require
320 modification; either to the methodologies, the associated assumptions or the
321 underlying data
- 322 • To test the accuracy of the default partition coefficients in predicting dissolved
323 metal concentrations.
- 324 • Identify any underlying uncertainties that may affect the performance in the
325 model
- 326 • Identify key improvements that might be made to the tool to improve the
327 predictive capability of the model at the National and catchment scale.

328

329 The SAGIS model has been developed for use at a waterbody and catchment spatial
330 scale and consequently has utilised nationally derived default assumptions where
331 measured data is not available, Finer resolution may be possible, but the accuracy of
332 outputs would need to be tested at a local level, potentially requiring more detailed
333 local data.

334

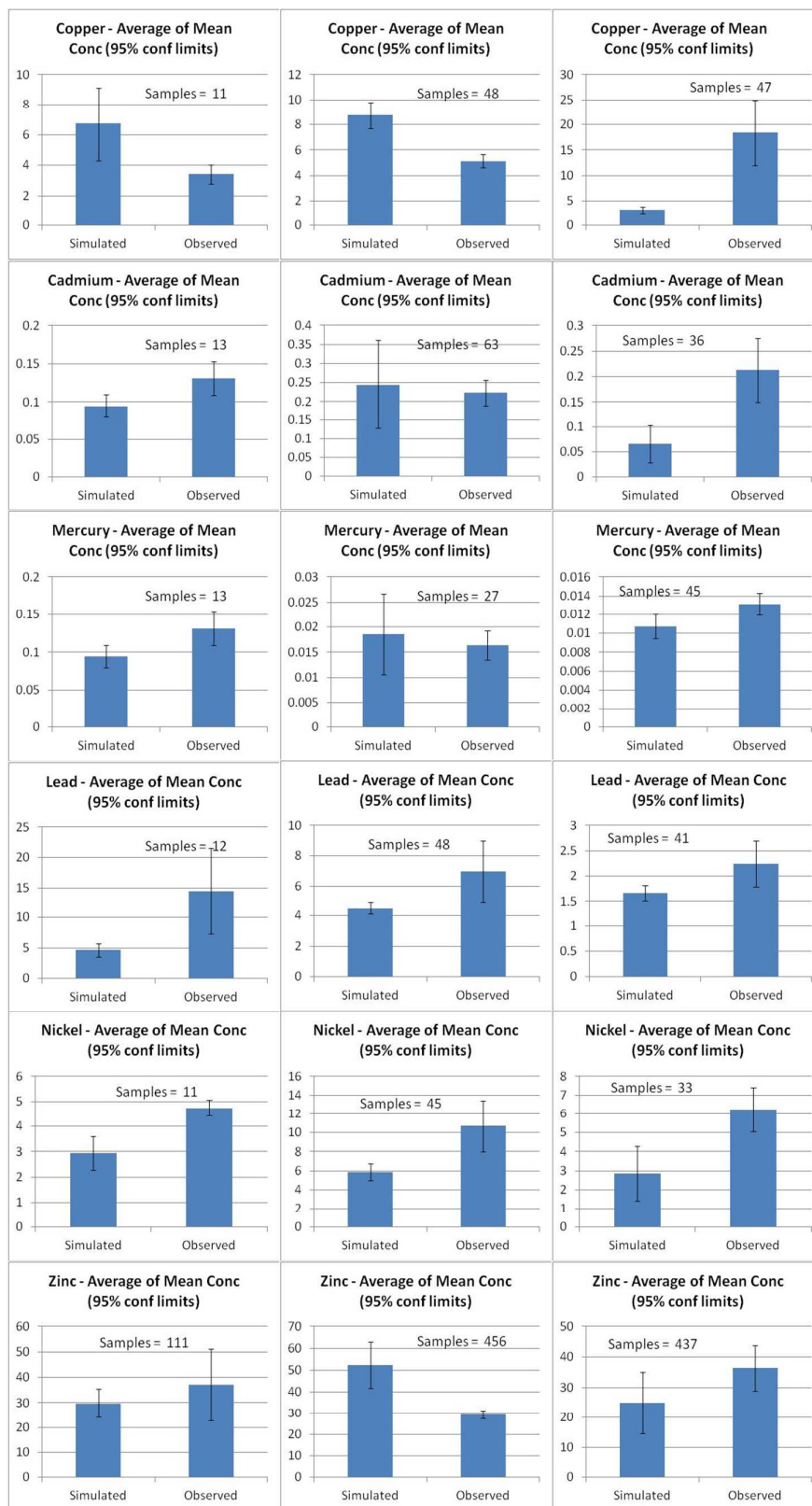
335 There are two key components of the model, the loads discharged to rivers,
336 waterbodies and catchments and their conversion into concentrations using the
337 SIMCAT water quality model. Although it would have been desirable to compare
338 measured and predicted loads entering the aquatic environment, there are no
339 national databases for measured loads for nay of the chemical parameters or input
340 sectors. There is however, an extensive national water quality monitoring database
341 held by the Environment Agency which was used for the purpose of testing the
342 SAGIS model. Furthermore, compliance testing is based on measurement of
343 concentrations against an EQS, and so it is vital that any model used by stakeholders
344 provides concentration data so that the impacts of any future measures applied to
345 improve water quality can be measured against EQS compliance.

346

347 Figures 4 and 5 provide summary statistics for the output of the SAGIS tool
348 compared with observed data (means derived from all monitoring between 2007 and
349 2009 inclusive) for each of the three regional model areas for metals and nutrients
350 respectively.

351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376

All concentrations as µg/l



377
378

Figure 4 Comparison between observed and simulated concentrations for all data generated from the three regional models for metals

379

380 The data in Figure 4 provide averages and 95% confidence intervals for a number of
381 metals for which concentrations can be simulated within the SAGIS programme. For
382 copper all predictions are of the same order, although simulations in the Northumbria
383 and Trent model tend to overestimate levels in the catchment slightly, with the
384 reverse for the SW model. It should be noted that for the SW model, observed values
385 were biased towards a number of highly mineralised sites in the west of the
386 catchment explaining the lower predicted value. Cadmium simulations were
387 comparable for Northumbrian and Trent regions but similar to copper, the model
388 underpredicted concentrations in the SW, for the same reasons. Mercury, being un-
389 influenced by UK mineralogy, shows good comparability between predicted and
390 observed means, although it should be noted that mercury monitoring data is
391 somewhat limited owing to many reported concentrations being less than limit of
392 detection. Lead concentrations, too are generally comparable, although in this case,
393 the NE mineralogy is dominated by lead/zinc mines and so as for Cu/Cd in the SW,
394 the model underpredicts lead inputs from this region. Concentrations in the Trent and
395 SW model are comparable and low. For nickel a metal for which anthropogenic
396 inputs dominate sources, a good comparison between measured and predicted
397 means is observed. Finally, for zinc, levels are relatively high and variable owing to
398 both the ubiquitous nature of zinc and its presence in soil, minerals and
399 anthropogenic discharges. However, the comparability is still good even for the mine
400 dominated sites in the Northumbrian model.

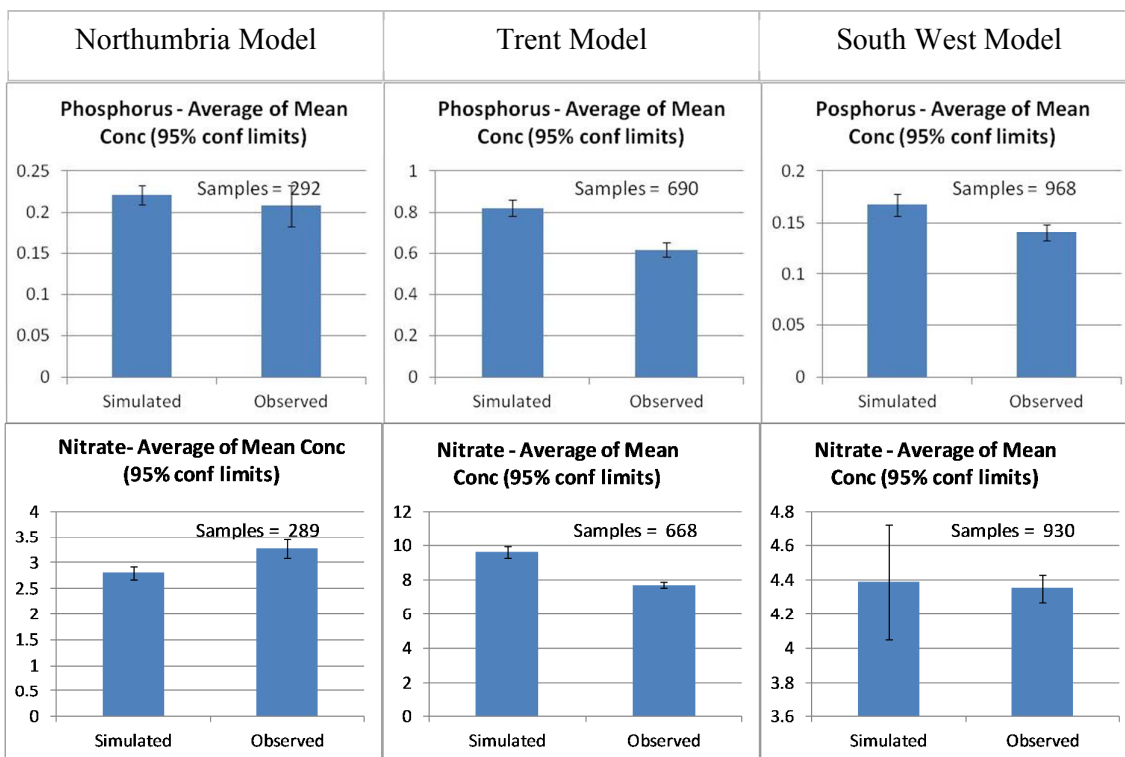
401

402 Overall it is clear from Figure 4 that given the nature of the generic datasets used
403 within the SAGIS model, comparisons with observed data may be considered very
404 good. Under estimates compared with measured data are associated with mineral-
405 rich areas where although some point source loads are accounted for, ie major adits
406 which are monitored for flow and concentration, thus providing a load, there are
407 many diffuse inputs not accounted for from minor adits, leaching and soil loss from
408 spoil tips and old processing sites. Previous studies in the Tamar catchment, for
409 example, have highlighted the loads of metals that can potentially arise from diffuse
410 sources, which may match or even exceed point source inputs.¹⁶ The model is
411 currently being updated to take account of diffuse mine inputs.

412

413 Data for nutrients (Figure 5) shows excellent agreement between predicted and
414 observed values for phosphorus and nitrate with significantly high levels in the urban

415 Trent region associated with discharges from wastewater treatment works (WwTW).
 416 Both of these inputs are dominated by a combination of agricultural diffuse and
 417 WwTW inputs. Given that agricultural inputs are derived from well developed and
 418 tested models and WwTW inputs are derived from extensive monitoring data, it is
 419 unsurprising that comparability is good. Furthermore, it needs to be noted that
 420 although observed data are collected all year round, sampling cannot take place
 421 during periods of heavy rain or flooding for safety and practicality reasons.
 422 Consequently, significant loads entering the waterbodies during high rainfall events
 423 particularly substances associated with particulate matter may be underestimated via
 424 routine monitoring data, thus leading to an underestimation compared with predicted
 425 data.
 426
 427



428
 429
 430 **Figure 5 Comparison between observed and simulated concentrations for**
 431 **all data generated from the three regional models for nutrients**

432
 433
 434

435 Although in some cases there appears to be a degree of statistically significant bias
436 they don't indicate a bias that is likely to be important in terms of status assessment
437 or catchment management. This suggests that either estimates of variance in the
438 model are too low or that there is bias in monitoring; or both.

439

440 The model was further validated by comparing measured versus predicted
441 concentrations within the individual rivers (Wear, Avon and Tame). Examples of
442 simulated concentrations of the case study substances in the main river channel are
443 provided in the Supporting Information (Figures 3, 4 and 5). There was insufficient
444 observed information for the organic substances to allow a meaningful comparison
445 between observed data and model output. For the river Wear in Northumbria zinc
446 inputs are dominated in the upper catchment by a limited number of mine water
447 inputs from abandoned zinc/lead mines. The model predicts accurately the step
448 change in concentration 25 km downstream of the source as major mine inputs enter
449 the river. After a small amount of dilution, levels then remain at ~ 32 µg/l to the tidal
450 limit. Comparison between predicted and measured values is excellent and further
451 supports the validity of the model outputs generated from summing a variety of point
452 and diffuse loads entering the river. Zinc levels in the river Tame are high owing to
453 inputs from historically contaminated land and resulting leachates (thought to be an
454 historic landfill site, EA data). Concentrations of several 100 µg/l are measured in the
455 upper catchment. These are not accurately predicted because no loads for inputs
456 from landfill or other contaminated land (with the exception of large mine sites) were
457 available on a nationwide basis. Further downstream, however, after substantial
458 dilution and where levels are influenced mainly from sewage effluent discharges (for
459 which adequate datasets are available), predictions match observed values. For the
460 agricultural catchment of the River Wylfe, levels of zinc are substantially lower and
461 predictions mostly lie within the 95% confidence intervals of measured data.

462

463 A similar exercise was carried out for phosphorus (Supporting Information, Figure 4).
464 The river Wear is subject to inputs from upland low intensity livestock farming and
465 low population centres. Overall predicted concentrations track observed values from
466 low concentrations of phosphorus in the upper catchment, slowly rising downstream
467 as larger towns provide phosphorus inputs to the river via sewage works. In all cases
468 except one anomalously high observed concentration predicted values were slightly
469 in excess of measured values, potentially owing to the PSYCHIC model over
470 estimating agricultural loads or a slight bias in the relationship between loads and
471 river flow used in the model. In a catchment such as the Tame, where accurate data

472 for the main inputs (sewage effluents) are available, predictions versus observed
473 values matched consistently down the catchment. A similar situation occurred for
474 phosphorus in the Wylve/Avon, a predominantly chalk catchment dominated by
475 arable farming, with good agreement between measured and predicted values
476 throughout the river length. Observed and predicted concentrations of phosphorus
477 show an increase at the 11km mark owing to a sewage treatment works input and
478 remain relatively consistent thereafter..

479

480 Simulated and observed nitrate concentration in the river Wear showed excellent
481 comparability with a gradual increase down the catchment as the contribution from
482 sewage effluents slowly increases. A similarly close agreement was observed for the
483 Tame, although much higher concentrations were observed and simulated (ca. 10
484 mg-N/l) owing to it being a sewage effluent-dominated catchment. The Wylve being a
485 more rural catchment exhibits lower concentrations which again are influenced by an
486 effluent discharge 11 km downstream of the source.

487

488 Overall, given the acceptable level of comparability between measured and predicted
489 concentrations it can be concluded that the model may be used for river planning
490 purposes with confidence, particularly when considering the impacts of applying
491 certain programmes of measures to meet environmental quality standards. This will
492 be of particular importance given mean catchment concentrations for Cd, Pb and Hg
493 will exceed the WFD EQS at certain sites with the potential of P, Zn and Cu also
494 exceeding UK derived limits in certain waterbodies. Cases where there is a
495 statistically significant bias between simulated and predicted data may require further
496 investigation at a local scale which is something regional Environment Agency staff
497 have been recommended to undertake.

498

499 **Discussion**

500

501 For the first time a model has been developed at a national scale which combines
502 predictions of input loads for all major sources of a contaminant with a water quality
503 model to predict in-stream concentrations for a number of determinands including
504 organics, metals and nutrients. Previously reported source apportionment models
505 have been developed at a catchment scale for single chemicals such as zinc¹¹ but
506 none has used national datasets to provide predictions over an area of approximately
507 150,000 Km² and including over 100 river catchments.

508

509 It is anticipated the framework could be used for:

- 510 • Water quality planning. SIMCAT has been widely used for water quality
511 planning and setting of wastewater consents. All of the existing functionality of
512 SIMCAT is retained in the tool and the diffuse and point source sector inputs
513 can be 'switched off' and the model used in the normal way if required. By
514 using the tool to estimate the contribution of the various diffuse and point
515 sources, this traditional use of SIMCAT can be enhanced by providing a better
516 indication on the sources of chemicals to improve planning of measures.
- 517 • Source control. SAGIS provides national and regional scale information on
518 source apportionment which can be used to inform national policy on source
519 control. For example, for organic chemicals the SAGIS could be used to
520 indicate whether the main sources of chemicals are from controllable sources
521 (e.g. rather than background).
- 522 • Reporting. SAGIS provides a range of visualisation options for chemical inputs
523 and predicted within-stream concentrations loads which could be readily used
524 for reporting of pressure characterisation and compliance for River Basin
525 Management plans.
- 526 • Testing of measures. SAGIS provides a framework to test the effectiveness of
527 measures related to each source sector and these can compared using the
528 model output and visualisation tools.
- 529 • Catchment management stakeholder engagement. SAGIS provides an
530 overview of the contribution of all sources of chemicals and, therefore, provides
531 the 'big picture' for a catchment to identify the dominant sectors and sources
532 and highlight where additional information is required or measures should be
533 targeted. Presentation of this overview using the visualisation functionality with
534 SAGIS provides a valuable starting point for stakeholder engagement through
535 provide the context at the catchment scale.
- 536 • Identify further monitoring and research. By bringing together a wide range of
537 national datasets, key areas of uncertainty in estimation of source
538 apportionment have become clearer which could provide a focus for improving
539 source data or the methodologies to create the export coefficient databases. By
540 showing which sectors are likely to be important for each chemical, the tool
541 provides a focus for where additional research and data collection would be
542 beneficial. Without this overview, this effort might be misdirected.

543

544 It is important to understand that the current version of the SAGIS is based on
545 national datasets and so lacks refinement at a local level. It is estimated that for

546 catchments in excess of 50 km² confidence in the model outputs can be considered
547 good. The model provides an open framework derived from the best available
548 national data and knowledge which may easily be refined at a local level.

549

550 No calibration or model conditioning has been carried out on the tool at present and
551 default values have been used in many cases; for example travel times and decay
552 rates. For assessment of compliance with water quality standards, data refinement,
553 calibration of decay rates and travel times (using the approach traditionally adopted
554 for SIMCAT) or conditioning of the export coefficients need to be carried out or the
555 assessment should take into account the difference between model output and
556 observed data.

557

558 The next step of its development is to provide it to local Environment Agency staff to
559 input data considered to be better than current default data currently in the
560 databases. Furthermore, additional substances are to be added over the coming
561 years and the databases maintained to ensure they are up to date. There will be a
562 particular focus on phosphorus in the next sets of river basin plans under the WFD
563 and so the interaction and data generated by PSYCHIC and used by SAGIS will be
564 more closely examined in the next phase of development, along with increasing the
565 coverage of the model from rivers to lakes, estuaries and coastal waters.

566

567 The SAGIS model represents the first comprehensive source apportionment tool to
568 be developed on a national scale for such a wide variety of chemicals. To meet ever
569 more stringent standards multiple interventions will be required to reduce discharges
570 from point and diffuse sources. SAGIS will assist regulators in making effective
571 decisions regarding how best to meet challenging water quality targets by identifying
572 the predominant source of a chemical. It also allows practitioners from sectors such
573 as the water industry to plan future improvements within a catchment at wastewater
574 treatment works where the greatest benefits on receiving water quality will be
575 achieved. This is of particular importance for ubiquitous substances such as nutrients
576 and metals where the impact of different sources can vary throughout a catchment.

577

578 SAGIS has been shown to provide accurate predictions of in-river concentrations for
579 metals and nutrients at a catchment scale, providing a degree of confidence in the
580 predicted outputs. The Environment Agency of England and Wales in conjunction
581 with key stakeholders will be using and developing the model as part of the second
582 cycle of river basin planning from 2013 onwards.

583

584 **Acknowledgements**

585

586 The authors would like to thank UKWIR, the Environment Agency including Mr Tony
587 Heaney, the Scottish Environmental Protection Agency and the Water Companies of
588 the UK for funding this project and supplying the data required to build the model.
589 Other organisations who have contributed data include the Highways Agency, Defra,
590 ADAS, CEH and FOREGS. We would also like to thank Dr Tony Warn for providing
591 an updated SIMCAT model and Mr Brian Ellor as UKWIR project manager and Karen
592 Murrell and Bob Crabtree from WRc for contributions associated with highway runoff.

593

594 **Supporting Information**

595 Tables of databases, default values and information used to develop the model are
596 provided in the supporting information in addition to comparison data for model
597 outputs versus observed data.

598

599 **References**

600

- 601 1) EU (2000) Directive 2000/60/EC of the European Parliament and of the
602 Council of 23 October 2000 establishing a framework for Community action in
603 the field of water policy.
- 604 2) EU (1976) The Directive 76/464/EEC of 4 May 1976 on pollution caused by
605 certain dangerous substances discharged into the aquatic environment of the
606 Community. Codified as 2006/11/EC.
- 607 3) EU (2008) Directive 2008/105/EC of the European Parliament and the
608 Council on environmental quality standards in the field of water policy,
609 amending and subsequently repealing Council Directives 82/176/EEC,
610 83/513/EEC, 84/156/EEC, 84/491/EEC, 86/280/EEC and amending Directive
611 2000/60/EC. Published in the Official Journal on 24 December 2008.
- 612 4) Lord, E.I. & Anthony, S.G. NEAP-N: A modeling framework for evaluating
613 nitrate losses at national and catchment scales. *Soil Use and Management*,
614 **2000**, 16, 167-174.
- 615 5) Davison, P. S., Withers, P. J. A., Lord, E. I., Betson, M. J. & Stromqvist, J.
616 PSYCHIC - A process-based model of phosphorus and sediment mobilisation
617 and delivery within agricultural catchments. Part 1: Model description and
618 parameterisation. *Journal of Hydrology*, **2008**, 350, 290-302.

- 619 6) Shanahan P., Henze M., Koncsos L., Rauch W., Reichert P., Somlyódy L.,
620 Vanrolleghem P. River water quality modeling: II. Problems of the art. *Water*
621 *Science and Technology*, **1998**, 38(11), 245–252.
- 622 7) Abbott M.B., Bathurst J.C., Cunge J.A., O'Connell P.E., and Rasmussen J.
623 An introduction to the European hydrological system: Systeme Hydrologique
624 European, 'SHE.' 2: Structure of a physically-based, distributed modeling
625 system. *Journal of Hydrology*, **1986**, 87, 61–77.
- 626 8) Jamieson D.G., and Fedra K. The 'WaterWare' decision-support system for
627 river-basin planning. 2: Planning capability. *Journal of Hydrology*, **1996**, 177–
628 198.
- 629 9) Kannel P.R., Kanel S.R., Lee S., Lee Y.S. and Gan T.Y. A Review of Public
630 Domain Water Quality Models for Simulating Dissolved Oxygen in Rivers and
631 Streams. *Environmental Modeling & Assessment*, **2011**, 16, 183-204.
- 632 10) SNIFFER (2003) Scottish and Northern Ireland Forum for Environmental
633 Research, Assessment of the development of a screening tool to identify and
634 characterise diffuse pollution pressures. WFD10, May 2003.
- 635 11) Hüffmeyer N., Klasmeier J. and Matthies M. Geo-referenced modeling of zinc
636 concentrations in the Ruhr river basin (Germany) using the model GREAT-
637 ER. *Science of the Total Environment*, 2009, 407, 7, 2296-2305.
- 638 12) UKWIR (2012) Chemical Source Apportionment under the WFD
639 (12/WW/02/3). Final report for UK Water Industry Research, 1 Queen Annes
640 Gate, London, ISBN: 1 84057 637 5.
- 641 13) FOREGS (2006) Geochemical Atlas of Europe. A contribution to IUGS/IAGC
642 Global Geochemical Baselines. R Salminen (chief editor) ISBN 951-690-913-
643 2.
- 644 14) UKSHS (2007) UK Soil and Herbage Pollutant Survey UKSHS Report No. 1.
645 Published by the Environment Agency, Rio House, Waterside Drive, Aztec
646 West, Almondsbury, Bristol, BS32 4UD, ISBN: 978-1-84432-692-1.
- 647 15) Rule K., Comber S., Ross D., Thornton A., Makropoulos C. and Ratui R.
648 Sources of priority substances entering an urban wastewater catchment –
649 trace metals. *Chemosphere*, **2006**, 63, 64-72
- 650 16) Minghanetara K., Braungardt C., Rieuwerts J. and Azizi F. Contaminant
651 fluxes from point and diffuse sources from abandoned mines in the River
652 Tamar catchment, UK. *Journal of Geochemical Exploration*, **2009**, 100, 116–
653 124.