

The ‘materials balance approach’ to pollution: its origin, implications and acceptance¹

Rüdiger Pethig, University of Siegen

January 2003

1. Introduction

Economic resources consist of services and natural resources (materials) taken from the environment and they are employed in production processes to generate consumer goods, both services and material products. After numerous transformations all materials are ultimately returned to the environment. Hence economics is not only about using scarce resources to satisfy consumers’ needs and wants – as all introductory economic textbooks tell us; economics is also about a large flow of materials (and energy) from the environment through various transformation processes, called production and consumption, back to the environment.

Conventional economic analysis always dealt with some aspects of that materials flow. Production functions link inputs to outputs assuming, often implicitly, that the materials embodied in the outputs are somehow transferred from the inputs to those outputs. Trade in materials and material products also captures a segment of the materials flow in that market clearing conditions secure the balance of material commodities given away and received. But on the whole, the economic materials flow is usually considered a rather unimportant and hence neglected side aspect of economic analysis that is accounted for only to the extent deemed necessary for explaining the production and consumption of positively valued (material) products.

The interest in the economic materials flow rose when in the 1960s cases of regional water and air pollution began to catch the attention of the public. For the first time spotlights were turned to the end of the materials flow, to some of the numerous residuals discharged into the environment. Selected residuals were quickly identified as the culprits of the environmental

¹ Helpful comments on an earlier version of this paper by Thomas Eichner and Kurt Schwabe are gratefully acknowledged. Remaining errors are the author’s sole responsibility.

degradation observed, and environmental policy began to react to emerging pollution hotspots like fire fighters do on fire alarm. Such an activist piece-meal approach is certainly unsatisfactory. In their seminal paper on “production, consumption and externalities” Ayres and Kneese (1969) (A&K, for short) reacted to partial ad hoc approaches by making a strong case for a comprehensive approach with a sound theoretical basis consisting of two parts: a systematic study of the residuals-generating materials flow with its link to pollution and the concept of pervasive (pollution) externality. Invoking the law of mass conservation and its corollary, the *mass balance principle* (MBP), they argued that the problem of environmental degradation cannot be adequately assessed unless the *complete economic materials flow* is envisioned with due regard of the MBP. The residuals at the end of the materials flow are determined by all materials entering that flow as well as by all transformation processes in production and consumption activities. The entire materials flow is therefore at the core of explaining environmental degradation whose feedback effects on the economy take the form of large-scale multi-party externalities due to missing markets.

A&K’s innovative contribution has become known as the ‘materials balance approach’ to pollution indicating that in essence it is a materials flow approach with strict regard of the MBP². Ayres and Kneese together with a small group of contemporary researchers, most of them associated with Resources for the Future, were aware that they had launched a research program on comprehensive analysis and management of residuals and pollution which was demanding and a challenge to conventional economic methodology. After some 30 years it is therefore appropriate and interesting to (i) recall the origin of this program, (ii) shed some light on its principal implications from today’s perspective and (iii) give a tentative assessment on how this research program has been received by the community of environmental and resource economists over the last decades. This is the aim of the present chapter. Of course, our assessment of A&K’s approach and its role in subsequent research will not be free from personal judgments, and the format of the present essay doesn’t allow reviewing the pertaining literature in an encompassing way.³

In section 2 we recall A&K’s ‘revitalization’ of the concept of externalities for the purpose of studying economy-environment interactions and argue that their externality approach has become the dominant environmental economics paradigm. Section 3 starts with observing that

² Following Dasgupta (1982, 162) we maintain that the ‘materials balance approach’ is not really *an approach* as such but rather an accounting device based on the law of mass conservation designed to ensure that economic activities are correctly described.

³ Among the major restrictions in scope are the neglect of applied work which ranked high in the pioneers’ priority list and the neglect of dynamic modeling.

much of the prevailing production theory does not adhere to the MBP. It reviews briefly how conventional production theory since A&K was amended to cope with joint production that is at the heart of the residuals problem. Then a production process with the option of abating production residuals is scrutinized in some detail. The consequences of imposing the MBP on this production technology turn out to be significant.

In Section 4 the production-cum-abatement technology of Section 3 is applied in a simple general equilibrium model where pollution negatively affects the utility of consumers and production technologies. The characterization of allocative efficiency reveals that and how much the MBP matters. Section 4 also offers a discussion along similar lines on the impact of the MBP in an economy with a production-cum-recycling technology. In this case two different kinds of materials are traced from ‘cradle to grave’ with the surprising result, among other things, that a product-design externality arises in addition to more conventional pollution externalities. Section 5 summarizes the main messages from MBP-regarding materials flow analysis and gives some tentative answers to the question of how environmental economists learnt the lessons from A&K’s ‘materials balance approach’.

2. "Production, consumption and externalities"

Some 30 years ago a small group of US economists published their innovative conceptual studies on the links between economic activities and pollution: A&K (1969), Kneese, Ayres and d’Arge (1970) and the associated body of contemporary work conducted innovative analyses of residuals and pollution laying the foundation, in retrospect, for an important part of what is now known and well-established as ‘environmental and resource economics’. These pioneers’ core contribution was the new concept of pervasive environmental externalities, based on a materials flow analysis that fully accounts for the MBP.

Their program was put forward clearly and forcefully in a seminal paper of Ayres and Kneese (A&K 1969) whose title we have chosen as the headline of the present section to highlight its path braking character. In a nutshell, its basic and at the time rather provocative message is that there are negative environmental externalities associated with the disposal of production and consumption residuals that are by no means exceptional, “... isolated and somewhat freakish aberrations” (Kneese, Ayres and d’Arge 1970, V) but rather normal, inherent, inevitable and pervasive. The present generation of young scholars probably doesn’t consider this message exciting at all, since in the meantime, externality theory has become and still is the

backbone of environmental economics (Pethig 2001). But for a full appreciation of the pioneers' merits it is helpful to recall that it was not before the late 1960s that cases of regional water or air pollution began to make their way into the headlines of the media. Economics as a science was caught by surprise and started to search for convincing explanations and problem solving strategies. To be sure, the concept of externalities was known since quite a while. Yet it was assigned a niche for non-market interactions of small numbers of parties (usually two!) and for rather irrelevant issues like bees and orchards. Allegorically speaking, A&K kissed the sleeping beauty 'externality' and placed it center stage for explaining and solving large-scale pollution problems.

Their innovative research program (to be discussed in more detail below) was not unanimously hailed in the academic community. There were those who insisted "... that the conventional framework and tools of economic theory are ill-adapted and in fact irrelevant for the analysis of ...[environmental disruption and social costs, R. P.]" (Kapp 1972, p. 95n.). Such scathing verdicts on the potential of economic theory to cope with pollution were perhaps too coarse to be convincing. A more refined attack was launched by economists who appeared to be uneasy with the claim of holding *pervasive market failure* responsible for the pollution observed. They perceived A&K's approach as a major challenge to the grand reconstruction of Adam Smith's invisible hand paradigm in rigorous neoclassical theory (Debreu 1959). Clearly, certain deviations from the ideal model were acknowledged, e.g. somewhat less than perfect competition, but they felt the thesis of large-scale market breakdown and subsequent serious misallocation needed to be rejected. The vivid dispute on this issue has become known as the Coase controversy relating essentially to the empirical relevance of welfare-reducing externalities: Coaseans contended that bargaining among rational self-interested parties of an externality would continue until all mutual gains from bargaining were exhausted, i.e. until all 'Pareto-relevant' externalities were eliminated. However, this conclusion presupposes well defined property rights and negligible transaction costs, a qualifier of the so called 'Coase theorem' that most environmental economists believed to be violated for pollution problems with large numbers of parties. In other words, Coasean economics doesn't predict efficient outcomes when many parties, common property resources and hence high transaction costs are involved. It rather emphasizes in such cases the relevance of an appropriate design of institutions and public policy for achieving allocative efficiency. Thus, somewhat ironically, the Coase theorem turned out to reinforce the externality-cum-inefficiency approach to pollution rather than weaken it.

In retrospect we observe that the revitalization and redirection toward large-scale pollution issues of the externality concept has been adopted and refined by the vast majority of environmental economists. The diffusion of this approach was quite fast, not only in America, but also across the Atlantic. Among the early European contributors and multipliers were Peter Bohm, Holger Bonus, F. R. Forsund, Karl-Göran Mäler, David Pearce and Horst Siebert. These authors and many others writing in the early 1970s explicitly linked their work to the A&K paper and to the related RFF research program, more generally. But after the first waves of perception and diffusion fewer environmental economists cited A&K in their work (Weinberg and Newbold, this volume). The fast and widespread adoption of A&K's externality approach by the international community of environmental economists probably amounted to 'voting by writing' more than to 'voting by citing'. Well established and broadly accepted professional methodology is usually not traced back to its origin each time it is applied.

There may be yet another reason for the profession's fading interest in A&K's pioneering paper. A&K made the MBP central to their analysis and over time, A&K (1969) was probably less perceived as dealing with pollution externalities than as advocating the 'materials balance approach' which many subsequent writers were reluctant to adopt. This development will be discussed in more detail below.

A&K's main attention is not focused on the *phenomenon* of pollution but rather on its *explanation*. No doubt, the identification of excessive discharges of residuals as the main culprit for pollution is a rather self-evident proposition and has probably been shared since long by many citizens concerned about the observed environmental degradation. However, A&K's principal innovative message supported by formal analysis is that to understand the dimension of and to cope with pollution properly one needs to look at the generation of residuals in the framework of a comprehensive materials flow analysis that is in line with the law of mass conservation. Hence the central role of their innovative and demanding study of the materials flow calls for a close look at the residuals issue and its treatment since A&K. In the following Section 3 we will therefore take up the analysis of residuals generation and materials flow and will return to the link between residuals and externality in Section 4.

3. Materials flow and production theory

Before A&K was published, neoclassical state-of-art general equilibrium theory ignored residuals altogether, not only the term but also the concept. This ignorance appeared to be ac-

ceptable so long as environmental degradation caused by residuals was negligible. But as mentioned before, the number of incidences of serious pollution increased in the 1960s. The discontent among environment-conscious economists grew about the conventional economic production technologies that were perceived as becoming increasingly at odds with the real world. While one is willing to agree that simple assumptions are desirable, in principle, one is reluctant to endorse simplicity taking the form of naivety (Sen 1985, 341).

3.1 A&K's materials flow analysis

To replace conventional production theory A&K suggested an innovative, comprehensive and consistent *materials flow approach* to the theory of production usually referred to as 'materials balance approach'. As is well-known the transformation of materials inputs into wanted and unwanted outputs is subject to the law of mass conservation.⁴ Material cannot be created from nothing and cannot disappear into the void. Essentially, the law imposes a material balance principle (MBP) on the entire flow of materials demanding the total weight of materials to be the same at all points along the materials flow. The special significance of the MBP for environment-economy issues is its implication that all material resources extracted from the environment, the 'source', eventually end up as residuals in the environment which functions now as a 'sink'. Most importantly, the total weight of material is unchanged on its route from source to sink. Thus one can infer the weight of residuals destined for the sink from the weight of materials taken from the source: "The amount of residuals inserted into the natural environment must be approximately equal to the weight of basic fuels, food, and raw materials entering the processing and production system, plus oxygen taken from the atmosphere" (A&K 1969, 284).

To model production subject to the MBP, A&K employ a static general equilibrium model with linear technologies, i.e. with fixed Leontief input-output coefficients, and partition the set of all productive factors into the subsets of tangible raw materials and services to trace the flow of materials from extraction through all sectors of the economy back into the environment. In their linear production model unwanted outputs (residuals) are generated in strict proportion to wanted outputs (*rigid joint production*). The amount of a residual can be reduced if and only if its by-product, the wanted good, is reduced, too, along with all other raw materials and services involved. The linear technology approach applied by A&K doesn't pose major problems for proper materials balance accounting, since wanted outputs as well as

residuals are generated in strict proportion to material inputs and other input services. Hence appropriate linear mass (or weight) balance constraints keep the analysis in line with MBP.

In the real world, technical relations between material inputs, residuals and wanted products are much more flexible, however, due to technological options such as factor and process substitution, recycling⁵ and waste abatement and treatment. Production and consumption technologies offer a variety of ways to transform raw materials into final goods and residuals and to transform final goods into residuals of different characteristics. Many alternative transformation processes are feasible that differ markedly in the chemical and physical composition of residuals they generate. This flexibility of transformation through factor substitution and flexible joint production is of utmost importance for environmental management, since different kinds of residuals usually differ in their detrimental environmental impact (to be discussed below in more detail) so that from a policy point of view one would like to push for changes of the residuals mix toward less harmful residuals even though their aggregate volume in terms of weight is unchanged owing to the MBP. By ignoring *available* transformation flexibilities one would likely overestimate the severity of pollution problems.

A&K are well aware of the limitations of their linear production model. They point out themselves that residuals production "... can be increased or decreased by investment, changes in material processing technology, raw materials substitution, and so forth" and that hence their "... assumption of unique coefficients ... is not consistent with the possibility of factor or process substitution or joint-production" (A&K 1969,298). For overcoming this deficiency they envision a production model with a very large collection of alternative sets of coefficients. Subsequent studies, notably Ayres (1972) and Russell and Spofford (1972) demonstrate that and how technological flexibility can be introduced in a linear framework allowing for "... choice among available alternatives for the production of goods and the transformation of the resulting residuals" (Russell and Spofford 1972, 138).

Far more general assumptions on production technology, including linear as well as non-linear production had already been employed in rigorous general equilibrium theory, such as Debreu (1959), but in the late 1960s this theory was hardly useful for anything but proving existence of a general competitive equilibrium by means of fixed point arguments. In contrast,

⁴ To simplify the exposition we ignore energy, set mass equal to material measured in terms of weight and normalize the units of all kinds of materials to be equal to one unit of weight.

⁵ A&K account for recycling in a 'linear' way but most of the subsequent literature on recycling didn't follow their approach.

input-output analysis with fixed technical coefficients allowed for tractable applied work yielding informative results, although it was a rather new analytical tool in the late 1960s too.

Given that state of the arts at the time, it was both well understandable and appropriate for A&K to follow the route of linear technologies. But during the next decades, innovative constructive algorithms for numerical computation of general equilibria and vastly improved electronic computing power have paved the way for large-scale computable general equilibrium (CGE) analysis capable to yield informative results. CGE models turned away from fixed-coefficients approaches to price-dependent coefficients (Conrad 2002) reflecting smooth non-linear production technologies that were most often applied in the analytical literature too. In recent years CGE studies have been conducted in ever increasing numbers in many different areas of economics, including environmental and resource economics. “The common approach in CGE modeling is .. to choose nested CES functional specifications which account for different degrees of substitutability between input factors on different nesting levels” (Conrad 2002, 91).⁶

Whether choice among different processes or smooth factor substitution is modeled does not seem to us a matter of principles but should rather be guided by pragmatic considerations. The point to be made is rather that the approach adopted needs to allow – approximately, at least – for the technological flexibility that can be identified empirically. The close relationship between both approaches is often demonstrated in microeconomic textbooks. Two or more different linear processes to produce one and the same commodity are shown to constitute a substitutional production function featuring convex isoquants with linear segments; and these isoquants become approximately smooth when the number of linear processes becomes sufficiently large.

In what follows we focus on smooth modeling of technological flexibilities not only because that is considered an appropriate approach by many researchers but also because bringing it in line with the MBP poses a greater challenge than activity analysis.

3.2 Materials flow, when production is non-linear

Factor substitution. The concept of smooth technological substitution has repeatedly been accused of not properly accounting for the MBP. Consider the Cobb-Douglas production function that is routinely employed in economic research and textbooks. With its isoquants

asymptotic to the axes this function allows for unlimited factor substitutability. Therefore it allows for asymptotic dematerialization of ‘material’ outputs which occurs, when the material inputs are successively substituted by non-material inputs combined, perhaps, with resource augmenting technical progress (Stiglitz 1974). But we usually think of material goods as containing a minimum amount of materials per unit. For such goods, the Cobb-Douglas function definitively fails to satisfy the MBP. But it can be reconciled with the MBP by bounding its domain, the input space, sufficiently away from the axes.

Just how wide or narrow substitution possibilities really are has triggered heated debates in the sustainability literature between adherents of weak and strong sustainability (Pezzey and Toman 2002). In that debate the dissent is mainly about the range of substitution possibilities between man-made and natural capital and therefore ultimately about the technical capacity to use substitution (and innovation) to offset environmental degradation. Clearly, the more limited substitution possibilities are, the greater is the impact of economic activities on the environment in the long term. Hence concerns about ecological-economic sustainability vary with the assessment of the substitution possibilities available in the real world. One should keep in mind, though, that the dispute is about empirical facts, and that reveals vast informational deficits about the nature of economic-ecological interdependences.

Joint production and residuals. Despite this dispute over facts, limitations in factor substitution don’t pose serious *theoretical* problems regarding its compatibility with the MBP. A more serious challenge is the role of residuals in conventional production theory. To grasp non-trivial residuals production one needs to invoke the theory of joint production that was already known and quite elaborate in the 1960s. Yet at that time its prime focus had not been on residuals and the environment but on the firm’s decision problem when marketable outputs were joint products. Two outputs are said to be joint products (by-products) when an increase in one output brought about by a suitable change in factor inputs is feasible, if and only if the other output is increased, too. Obviously, for environmental economics joint production of wanted outputs (goods) and unwanted outputs (residuals) is of particular interest. This kind of joint production is not an exceptional case but rather pervasive and inevitable. For example, when material consumer goods reach the end of their useful economic life they are bound to

⁶ Nevertheless, some methodological plurality can still be observed. For example, Faber et al. (1999) adhere to linear technologies allowing for discrete and costly process substitution in an intertemporal framework.

turn into residuals. Residuals also emerge inevitably during the production process of turning material inputs into wanted outputs.⁷

Debreu's (1959) axiomatic production theory is general enough to include a wealth of joint production. Yet in this model's general equilibrium there is no excess supply of products⁸ destined to be discharged into the environment. What appears as a contradiction at first glance is resolved by observing that the residuals problem is simply 'assumed away' via the axiom of 'free disposal'⁹. It implies that if some firm generates an unwanted product you will always find another firm which is capable and willing to take over, at zero price, any non-negative amount of that product as an input; it is capable to do so, because due to the assumption of free disposal its production remains feasible with that unwanted product as an additional input and it is willing, because its profit-maximizing production plan is not affected by an additional zero-priced input. As a consequence, in Debreu's production world there do not exist 'unwanted' products that are generated as by-products of wanted products. All products are at least 'weakly wanted' (including all toxic wastes!) by some firm as a consequence of the free disposal axiom. This axiom is certainly at odds with the real world.

It took almost 20 years before Bergstrom (1976) proved that one can drop the free disposal assumption without being forced to compromise on the generality of the remaining axioms. To assess the consequences of the absence of free disposal imagine first that there exist competitive markets for *all* outputs irrespective of whether they may be wanted or not. Then the market clearing price of some commodities would turn out to be negative indicating that those commodities are 'bads'.¹⁰ Since demand equals supply for all these bads, no residuals would be released into the environment at all. No doubt, this is a purely virtual world, first, because production with zero discharge of residuals is not feasible¹¹, and second, because negative

⁷ For more details on joint production and its link to thermodynamics see Faber et al. (1998) and Baumgärtner (2002). For the purpose of the present paper it suffices to invoke the simple 'folk theorem' which has it that the transformation of a single material necessarily yields more than one output.

⁸ More precisely, a necessary condition for a product to be in excess supply in general equilibrium is that its equilibrium price is zero (free good). But due to 'free disposal' (to be discussed below) there always exists another equilibrium where all zero-priced products exhibit zero excess supply.

⁹ Free disposal means that production is feasible with any amount of inputs of any kind (including residuals). Technically speaking, the entire non-negative orthant of the commodity space is a subset of each firm's production set.

¹⁰ It is worth mentioning that the subsets of goods and bads (residuals) are endogenously determined in general equilibrium (Mäler 1974, Pethig 1979). In such an approach Baumgärtner (2000) calls joint production 'ambivalent' since one doesn't know in advance whether a by-product will turn out to be a good or a bad in equilibrium. As an example, waste paper may have a positive market price if its demand for recycling is sufficiently strong but its price is negative if people have to pay for its (mandatory) removal. Positively priced waste paper is not a residual for discharge but rather a valuable intermediate product for some downstream firm.

¹¹ Consider, for example, the case of burning fossil fuels. One would have to collect all CO_2 -gas generated in containers and hand it then over to firms which somehow prevent that gas from entering the environment.

prices cannot prevail in an unregulated economy with an option to costlessly discharge bads into the environment. In other words, without effective enforcement of negative prices bads will be zero priced and their excess supply will be dumped into the environment. If negative prices *are* enforced, they constitute emission taxes, in fact, and increasing tax rates (in absolute value) can be expected to reduce the excess supply of bads alias residuals (Pethig 1979).

Although the rigorous theoretical development reported here took place after A&K, it is amazing how well Ayres and Kneese already understood the pertaining theoretical issue. They mention the possibility of introducing negative (virtual) prices in general equilibrium theory and also point out the “... underlying similarity of negative prices and effluent taxes” (A&K 1969, 292, fn.24).

3.3. *Production, waste abatement and the materials balance principle (MBP)*

Flexible joint production with incomplete materials flow. Though Debreu-style set-theoretic production technology without ‘free disposal’ is intellectually appealing it is too general to yield informative results. Therefore, most writers in environmental economics employed differentiable production functions that may be quite general, too, as e.g. in Baumol and Oates (1975) or very simple as in Pethig (1976). Many other studies have followed this route. To describe and explain environmental problems in a non-trivial way differentiable production functions must have a core characteristic: they need to deal with joint production of wanted and unwanted products. To fix our ideas, suppose now there is a production activity that produces two outputs, a wanted output and an unwanted (and environmentally harmful) one, with a given set of inputs whose quantities are fixed. The resultant mapping in the output space is well known as transformation curve. Quite obviously, a necessary condition for a non-trivial pollution externality problem to arise is that the transformation curve is positively sloped¹²: society can choose from a menu of good and bad things but to get more of the good it needs to swallow more of the bad.

To illustrate this concept of *flexible joint production* in the simplest possible way consider the concave (and differentiable) production function

$$y = Y(\underset{+}{e}, \underset{+}{\ell}, \underset{+}{m}), \tag{1}$$

¹² Here we consider only what we call flexible joint consumption. In case of rigid joint production, e.g. with linear technologies, and fixed amounts of all inputs the transformation curve would degenerate to a single point in output space.

where the plus signs in (1) mean that the respective partial derivative is positive, i.e.¹³ $Y_e > 0$, $Y_\ell > 0$ and $Y_m > 0$. In (1), y is the amount of a consumer good produced with some material m and with labor ℓ . By e (with e for emissions) we denote the amount of a residual generated in the process of producing y . Note that e is an output just like y although it is formally treated as an input. In fact, (1) characterizes e as a flexible by-product of y , since for constant ℓ and m (1) yields a positively sloped and concave transformation function. Technologies such as (1) continue to be routinely applied but authors usually don't pay any attention to the materials flow implications. To check the compatibility of (1) with the MBP let us assume the output y and the quantities m and e are all measured in terms of weight (see footnote 2). Suppose further (1) describes a transformation process in which m is completely transformed into the outputs y and e such that $m = e + y$ holds. Obviously, this equation is compatible with $y = Y(e, \ell, m)$, if and only if $Y_e = -1$, contradicting our assumption $Y_e > 0$. The striking – and sad – conclusion is that (1) doesn't represent a consistent modeling of the implied materials flow, it doesn't trace and record all relevant flows of materials from inputs to outputs.

We end up with the intriguing observation that (1) can be credited for taking up the concept of joint production of wanted and unwanted products in a non-trivial way but is not in line with the MBP, when (1) is viewed as a process of transforming the material input m into the material outputs y and e . In what follows we will demonstrate that the mapping of the materials flow as described in (1) is incomplete by showing that (1) is consistent with an MBP-regarding production-cum-abatement technology but represents that technology in a 'truncated' form only.

Production-cum-abatement with complete materials flow. To specify such a technology, consider the technical relations:

$$y = F \left(\underset{+}{\ell}, \underset{+}{m_y} \right) \quad (2a) \quad a = H \left(\underset{+}{m_a} \right) \text{ with } H(0) = 0 \quad (2e)$$

$$r_y = m_y - y \quad (2b) \quad e := r_y - a \geq 0 \quad (2f)$$

$$r_{a_1} = a \quad (2c) \quad m_a + m_y =: m \quad (2g)$$

$$r_{a_2} = m_a \quad (2d) \quad r_c = y \quad (2h)$$

¹³ When upper case letters represent functions, subscripts attached to those letters indicate partial derivatives.

The letters e , ℓ , m and y have the same meaning as in (1). r_{a_1} , r_{a_2} , r_y and r_c are different kinds of residuals. r_y is the amount of a residual generated in the course of producing y (called y -residual). r_{a_1} and r_{a_2} are residuals, called abatement residuals, generated in the process (2c) – (2f) of abating y -residuals. Since waste abatement does “... not destroy the residuals but only alter their form” (A&K 1969, 283) abatement residuals are bound to emerge in production-cum-abatement technologies. Hence they have got to be explicitly accounted for. The abatement activity as modeled in (2) uses material m_a and the amount of y -residuals, a , to transform a and m_a into the residuals r_{a_1} and r_{a_2} as determined by (2c) and (2d). Equation (2e) entails the information that it is technically feasible to use the material input m_a for transforming the amount a of y -residuals into a_1 -residuals (2c). However, m_a does not vanish but is turned into a_2 -residuals (2d) when the abatement process is finished. $e := r_y - a$ from (2f) - with the letter e for emissions - is the amount of y -residuals left over after abatement and destined for discharge into the environment. r_c is the amount of residuals emerging at the end of the useful life of the consumer good, called post-consumption residuals, for short. (2a) and (2b) model y and r_y as joint products, i.e. as two outputs generated by the same set of inputs (ℓ, m_y) . Through increasing the labor input it is feasible, for any given m_y , to shift the output mix toward y . But the constraints on technological feasibility implied by the material balance equation $r_y + y = m_y$ from (2a) are by no means trivial: Due to the second law of thermodynamics, $r_y > 0$ and $\partial r_y / \partial m_y \in]0, 1[$ hold, and for these reasons (i) the domain of F must be constrained to all (ℓ, m_y) satisfying $F(\ell, m_y) < m_y$ and (ii) we require $\partial r_y / \partial m_y = 1 - F_m \in]0, 1[$. The latter property is why Cobb-Douglas is ruled out as a functional form for F . (2h) expresses the assumption that the mass of y is not destroyed through consumption. The weight of consumption goods fully carries over to the consumption residuals, r_c .

In the present context our concern is not whether (2) precisely describes an empirically observed abatement technology though the specifications (2c) - (2e) don't seem to be implausible. They are primarily chosen for convenience of gaining conceptual insights. What is important in that regard is that (2) adheres to the MBP and thus describes a production-cum-abatement process that is no more and no less than a transformation process. The (only) material input m is transformed in (2) into the ‘ultimate outputs’ e , r_{a_1} , r_{a_2} and r_c . Together these

four outputs absorb all ‘transformed’ material input, m , and don’t contain any substance not stemming from m . Since all material inputs and outputs are measured in terms of weight, the equations (2) take precisely care of the law of mass conservation. It is easy to check that this transformation satisfies the MBP: $e + r_{a_1} + r_{a_2} + r_c = m$.

Suppose for a moment, the function F is Leontief and H is linear. Then we would be back, essentially, in A&K’s MBP-regarding materials flow analysis. We will not pursue here any further that kind of analysis because of its unrealistic feature of abatement being an all-or-none activity.¹⁴ Instead, we now aim at clarifying the link between conventional non-linear production theory as exemplified by (1) and its MBP-regarding non-linear (and differentiable) counterpart (2).

In view of (2b), (2e), (2f) and (2g) one obtains

$$m_y = e + H(m - m_y) + F(\ell, m_y). \quad (3)$$

Applying the implicit function theorem we infer from (3) that there is a function M satisfying

$$m_y = M(e, \ell, m) \quad \text{with} \quad M_e = \frac{1}{A_o} > 0, \quad M_\ell = \frac{F_\ell}{A_o} > 0 \quad \text{and} \quad M_m = \frac{H_m}{A_o} > 0, \quad (4)$$

where $A_o := 1 + H_m - F_m > 0$. Next we insert (4) into (2a) to get

$$y = F[M(e, \ell, m), \ell] =: \tilde{Y}(e, \ell, m). \quad (5)$$

which yields immediately $\tilde{Y}_e := F_m M_e > 0$, $\tilde{Y}_\ell := F_m M_\ell + F_\ell > 0$, and $\tilde{Y}_m := F_m M_m > 0$.

Since the production functions Y from (1) and \tilde{Y} from (5) are qualitatively the same (after having added some conditions to secure concavity of \tilde{Y}), *equation (1) turns out to be a truncated version of technology (2)*. (1) as well as (5) are ‘truncated’ in the sense that the proper technology (2) describes the production of one intermediate output, y , and four ultimate outputs, namely r_c, e, r_{a_1} and r_{a_2} , with two inputs, ℓ and m whereas (1) and (5), respectively, capture only the intermediate output, y , and the final output e .

Taking into account that (5) is based on (2) closer inspection reveals a not so obvious difference between (1) and (5) which highlights the relevance of the MBP. Suppose for a moment, $Y_e > 0$ for all $e \geq 0$. It would then be possible for any given inputs (ℓ, m) to arbitrarily expand the output y by increasing e – in outright violation of the MBP. An appeal to common sense

¹⁴ If any technological assumption was ever overwhelmingly embraced by environmental economists in both theoretical and applied work, it is certainly the notion of increasing (as opposed to constant) marginal abatement costs. The equivalent in our model is H being increasing and strictly concave in m_a .

suffices to concede that there has got to be an upper bound on expanding the materials outputs y and e when m is given. To make that proposition precise combine (2a), (2b), (2e) and (2f) to obtain $e = m_y - F(\ell, m_y) - H(m - m_y)$ with $\partial e / \partial m_y = 1 - F_m + H_m > 0$. Hence e is strictly increasing in m_y (with ℓ and m constant), and therefore

$$e = m - F(\ell, m) \quad \text{or, equivalently,} \quad m = m_y = M(e, \ell, m) \quad (6)$$

defines the maximum amount of e that can be generated when the inputs are (ℓ, m) . With (6) placing an upper bound for e when ℓ and m are given, the domain of the function \tilde{Y} is not the entire three-dimensional non-negative orthant but rather a proper subset of it, namely $\{(e, \ell, m) \geq 0 \mid e \leq m - F(\ell, m)\} =: \tilde{D}$. Studies employing production functions of the type (1) without regard of (2) usually do not explicitly account for the proper domain constraint of the production function. Instead, they assume $Y_e = 0$ for large e , as e.g. Tahvonen and Kuuluvainen (1993). While a blatant contradiction to the MBP can be avoided in this way, $Y_e = 0$ is clearly an ad hoc assumption that is not in line with (2). To see that we maximize the function \tilde{Y} with respect to e subject to the constraint $e \leq m - F(\ell, m)$. Since this constraint is strictly binding at the maximum, the Lagrange multiplier associated with it, say μ , is strictly positive. Hence $\tilde{Y}_e = \mu > 0$. In other words, the function \tilde{Y} exhibits $\tilde{Y}_e > 0$ on its entire domain \tilde{D} with all its boundary points $(e, \ell, m) \in \{(e, \ell, m) \in \tilde{D} \mid e = m - F(\ell, m)\}$. We conclude that the ad hoc assumption $Y_e = 0$ in studies using (1) is not compatible with the MBP, if (1) is meant to be a truncated form of (2).

According to (2g) the factor m is either employed to produce the consumer good, m_y , or to abate y -residuals, m_a . (2) implies that for given (ℓ, m) the amount of material, m_a , completely determines the quantities of all outputs:

$$y = r_c = F(\ell, m - m_a), \quad e = m - m_a - F(\ell, m - m_a) - H(m_a), \quad r_{a_1} = H(m_a) \quad \text{and} \quad r_{a_2} = m_a, \quad \text{with}$$

$$\frac{\partial y}{\partial m_a} = \frac{\partial r_c}{\partial m_a} = -F_{m_y} < 0, \quad \frac{\partial e}{\partial m_a} = -(1 - F_m) < 0, \quad \frac{dr_{a_1}}{dm_a} = H_{m_a} > 0, \quad \text{and} \quad \frac{dr_{a_2}}{dm_a} = 1 > 0.$$

These derivatives show the impacts on all outputs of shifting material from production to abatement. Increasing abatement, or more precisely, $dm_a = -dm_y > 0$, reduces the amount of

both y -residuals ($\partial e / \partial m_a < 0$) and the consumer good ($\partial y / \partial m_a < 0$). It is exactly this effect that is captured in $\tilde{Y}_e > 0$. But the other derivatives above demonstrate that stepping up abatement has also an impact on each of the three remaining residuals: it curbs post-consumption residuals and expands the generation of abatement residuals. These side effects are missing in (1) as well as in (5). It is possible, of course, to ignore the changes in r_c , r_{a_1} and r_{a_2} induced by variations in the net amount of y -residuals, e . However, given the technology (2) with its accurate mass balance accounting, one cannot claim that these changes do not occur. They are inadvertently and inevitably linked to expanding abatement.

These findings suggest immediately that it depends on the impact on pollution of each kind of residuals contained in (2) to assess how appropriate it is to rely on (1) instead of (2). We will take up again and investigate more thoroughly this central issue in the following Section. But before doing so it is useful to briefly characterize A&K's approach to environment-economy interdependence.

4. Materials flow and pollution externalities

4.1 A&K's modeling of environment-economy interdependence.

To identify the links between residuals and externalities it is necessary to treat residuals and externalities as two entirely distinct concepts. *Residuals* are materials generated in production and consumption activities that have no positive value for any economic agent.¹⁵ In contrast, an *externality* is present (Baumol and Oates 1988, p. 17) "... whenever some individual's ... *utility* or *production* relationships include real ... variables, whose values are chosen by others ... without particular attention to the effects on ... [that individual, R.P.]"

In general, there is no simple, let alone linear relation between residuals and environmental externalities. Residuals discharged into the environment have the *potential* to change the state of the environment depending on the types and quantities of residuals emitted, on the locations of discharge and on the assimilative capacity of the environmental media at issue which in turn depends on past emissions. The link between residuals emitted and the state of the environment can be described by environmental diffusion models which according to Russell and Spofford (1972, p. 122) may be essentially thought of "... as transformation functions

operating on a vector of residuals discharges and yielding another vector of ambient concentrations at grid points throughout the environment". Residuals causing environmental degradation with negative feedback effects on (other) economic agents are usually called 'pollutants'.

A&K's (1969, 291) still modern view is that the environment offers economic agents¹⁶ (i) common property resources, (ii) services to assimilate residuals and (iii) non-rejectable pollutants (as disservices), and all associated exchanges occur without the moderation of markets. The latter two services are closely linked, of course. The environment's ability to accept, dilute, chemically degrade and neutralize residuals, its so called assimilative capacity, is limited. When it is exceeded, residuals discharged into the environment cause ambient pollution that, in turn, adversely affects economic agents' utility or production functions. The pollution is imposed on a large number, or even on all, individuals and those who release emissions pay no particular attention to the effects of that pollution on others. Hence an externality arises in the sense of the Baumol/Oates definition.

A&K provide a formal analysis of the link between residuals, pollution and feedback effects. They extend their input-output framework characterized by fixed, unique coefficients (excluding process substitution and abatement) to account for environmental services and disservices. Yet their activity analysis makes all these services linear and turns residuals into pollutants in strict proportion. With great awareness of the limitations of their approach they provide an extensive discussion on modifications that would be necessary and desirable to capture the formation of pollution and its feedback effects on the economy in a more satisfactory way. For this reason and because subsequent developments opened a more convincing route of modeling the link between residuals and pollution, we refrain from discussing A&K's analysis further here. Yet it is worth mentioning that A&K (1969, fn. 23) hint to the theory of public goods as a potentially promising framework to model the links between residuals, pollution and its feedback effects. In subsequent work, the theory of public goods has, in fact, become standard for modeling pollution externalities.

In what follows, we briefly outline a simple public goods model of pollution containing the technologies (1) and (2), respectively. Our main intention is to answer the question what the

¹⁵ In view of our remarks in footnote 10 one would have to distinguish between 'residuals for further use' and residuals for disposal' (or between wanted and unwanted residuals). But to keep the exposition simple we don't apply that terminology.

¹⁶ A&K consider producers only but the impact of the disservice of non-rejectable pollution on consumers is also severe and has thoroughly been investigated during the last decades.

role of the MBP-regarding materials flow is when the link between residuals, pollution and its feedback effects is explicitly modeled.

4.2 Production, abatement and pollution in a simple general equilibrium model

Characterization of allocative efficiency. Denote by p an index of the ambient concentration of pollutants, called pollution for short, and define

$$p = P \left(\underset{+,0}{e}, \underset{+,0}{r_{a_1}}, \underset{+,0}{r_{a_2}}, \underset{+,0}{r_c} \right) \quad \text{with } P(0, 0, 0, 0) = 0 \quad (7)$$

as the pollution that results from releasing the residuals e , r_{a_1} , r_{a_2} and r_c into the environment. Clearly, $P_x \geq 0$ for $x = e, r_{a_1}, r_{a_2}, r_c$, and we expect $P_x = 0$ to hold if and only if the quantity of the residual under consideration is small enough to be fully neutralized by nature's assimilative capacity.¹⁷ Environmental degradation as manifested in p is now assumed to affect both consumers' utility and the conditions to produce the consumer good¹⁸:

$$u_i = U^i(y_i, p) \quad i = 1, 2, \dots, n \quad (8)$$

$$y = \hat{Y}(e, \ell, m, p) := \underset{-}{G}(p) \cdot \underset{+}{\tilde{Y}}(\underset{+}{e}, \underset{+}{\ell}, \underset{+}{m}) \quad (9)$$

Our simple general equilibrium model is completed by introducing the standard resource constraints

$$y \geq \sum_i y_i, \quad \bar{\ell} \geq \ell, \quad \text{and} \quad \bar{m} \geq m, \quad (10)$$

where $\bar{\ell}$ and \bar{m} are the economy's fixed factor endowments. The next step is to characterize a Pareto efficient allocation of the economy given by (2) and (7) – (10). For that purpose consider the Lagrangean

$$\begin{aligned} L = & \sum_i \alpha_i U^i(y_i, p) + \lambda_f [\hat{Y}(e, \ell, m, p) - y] + \lambda_p [p - P(e, r_{a_1}, r_{a_2}, r_c)] + \lambda_y (y - \sum_i y_i) + \\ & + \rho_{a_1} [r_{a_1} - H[m - M(e, \ell, m)]] + \rho_{a_2} [r_{a_2} - m + M(e, \ell, m)] + \rho_c (r_c - y) + \\ & + \lambda_\ell (\bar{\ell} - \ell) + \lambda_m (\bar{m} - m) + \lambda_m [m - M(e, \ell, m)], \end{aligned} \quad (11)$$

¹⁷ We acknowledge the severe limitations inherent in static approaches to assimilative capacity and pollution. Many studies show how important the dynamics of economy-environment interactions are (e.g. Perrings 1986, 2001, van den Bergh and Nijkamp 1994). However, the economy itself is also a dynamic system and yet, static analysis has proved to be an adequate method to tackle many (though not all!) important issues.

¹⁸ A&K had addressed the negative productivity effects of pollution. To incorporate these effects into our model, we extend our technology (2) as shown in (9). In (9), the impact of p on production has been given the rather restrictive form $G(\cdot) \cdot \tilde{Y}(\cdot)$ in order to link the present model firmly to the technology (2) with $\tilde{Y}(\cdot)$ as defined in (5). Obviously, the productivity effect of pollution can easily be 'switched off' by setting $G_p(p) \equiv 1$.

where α_i for $i = 1, \dots, n$ are arbitrary positive numbers. Note that the equations (2a), (2b) and (2f) are not considered in (11), because the information they carry is already fully contained in $\hat{Y}(e, \ell, m_a + m_y, p)$ from (9). Focusing on an interior solution¹⁹ of (11) the relevant first-order conditions are

$$\alpha_i U_y^i = \lambda_y, \quad i = 1, \dots, n \quad (12a) \qquad \lambda_y = \lambda_f + \rho_c \quad (12f)$$

$$\lambda_p = - \left(\sum_i \alpha_i U_p^i + \lambda_f \hat{Y}_p \right) \quad (12b) \qquad \rho_{a_1} = \lambda_p P_{r_{a_1}} \quad (12g)$$

$$\lambda_p P_e = \lambda_f \hat{Y}_e + (H_m \rho_{a_1} + \rho_{a_2}) M_e \quad (12c) \qquad \rho_{a_2} = \lambda_p P_{r_{a_2}} \quad (12h)$$

$$\lambda_f \hat{Y}_\ell = \lambda_\ell - (H_m \rho_{a_1} + \rho_{a_2}) M_\ell \quad (12d) \qquad \rho_c = \lambda_p P_{r_c} \quad (12i)$$

$$\lambda_f \hat{Y}_m = \lambda_m + (H_m \rho_{a_1} + \rho_{a_2}) (1 - M_m) \quad (12e)$$

We now condense the information contained in (12). As a first step, we invoke (12f) and (12i) to obtain $\frac{\lambda_y}{\lambda_f} = \frac{\lambda_f + \rho_c}{\lambda_f} = 1 + \frac{\lambda_p}{\lambda_f} P_{r_c}$. Next we combine this equation with (12a) to turn (12b)

into

$$\frac{\lambda_p}{\lambda_f} = - \left(\frac{\lambda_y}{\lambda_f} \sum_i \frac{U_p^i}{U_y^i} + \hat{Y}_p \right) = - \left[\left(1 + \frac{\lambda_p}{\lambda_f} P_{r_c} \right) \sum_i \frac{U_p^i}{U_y^i} + \hat{Y}_p \right] = \frac{- \left(\sum_i \frac{U_p^i}{U_y^i} + \hat{Y}_p \right)}{1 + P_{r_c} \sum_i \frac{U_p^i}{U_y^i}} \quad (13)$$

In (13), $\left(1 + P_{r_c} \sum_i \frac{U_p^i}{U_y^i} \right) \in]0, 1[$ since $\lambda_p / \lambda_f > 0$ by assumption (see footnote 19). Observe

also that (12c), (12h) and (12i) imply $\frac{\lambda_p}{\lambda_f} = \frac{\hat{Y}_e}{P_e - (H_m P_{r_{a_1}} + P_{r_{a_2}}) M_e}$ which turns (13) into our

central result

$$MP_p = \frac{\hat{Y}_e}{P_e} \cdot Q, \quad \text{with } MD_p := - \left(\sum_i \frac{U_p^i}{U_y^i} + \hat{Y}_p \right) \text{ and } Q := \frac{\left(1 + P_{r_c} \sum_i \frac{U_p^i}{U_y^i} \right) P_e}{P_e - (H_m P_{r_{a_1}} + P_{r_{a_2}}) M_e}. \quad (14)$$

¹⁹ An interior solution implies, in particular, that it is optimal to abate y -residuals ($a > 0$). Therefore $\lambda_m = 0$. Observe also that all economic variables are non-negative and all terms following the Lagrange multipliers in (11) are non-negative and concave. Due to the Kuhn-Tucker theorem all Lagrange multipliers are therefore non-negative in the solution of (11). Unless stated otherwise we consider strictly positive solution values of all Lagrange multipliers except for λ_m .

Clearly, (14) characterizes the efficient level of pollution from all residuals discharged into the environment. MD_p is the marginal pollution damage, and hence the marginal benefit of pollution control, consisting of two external effects, (i) the negative effect of pollution on the production of the consumer good, $|\hat{Y}_p|$, and (ii) the impact of pollution on consumers measured by the consumers' aggregate marginal willingness to pay (in terms of the consumer good) for avoiding an increase in pollution, $|\sum_i (U_p^i / U_y^i)|$. The right side of (14) represents the marginal benefit of pollution [or equivalently, the marginal cost of pollution control] in terms of the increment of the consumer good gained through pollution [or lost through pollution control]. $\hat{Y}_e Q / P_e$ is the marginal cost of reducing the emission of *all* residuals, while \hat{Y}_e is the marginal cost of abating *y*-residuals. If we multiply both sides of (14) with P_e we find that $MP_p \cdot P_e = \hat{Y}_e \cdot Q$, where $MP_p \cdot P_e$ is the marginal environmental damage of *y*-emissions.

It follows from (14) and the definition of Q that in case of $P_e > 0$ it is true that

$$MP_p \cdot P_e \begin{cases} > \\ = \\ < \end{cases} \hat{Y}_e \quad \text{if} \quad \begin{cases} P_{r_{a1}} > 0 \text{ and/or } P_{r_{a2}} > 0 \text{ and } P_{r_c} = 0, \\ P_{r_{a1}} = P_{r_{a2}} = P_{r_c} = 0, \\ P_{r_{a1}} = P_{r_{a2}} = 0 \text{ and } P_{r_c} > 0. \end{cases} \quad (15)$$

The striking result is that *the optimal control of y-emissions depends on the marginal impact on pollution of the emission of all other residuals*, i.e. on the magnitudes of $P_{r_{a1}}$, $P_{r_{a2}}$ and P_{r_c} . Suppose first, pollution is caused by *y*-emissions exclusively ($P_{r_{a1}} = P_{r_{a2}} = P_{r_c} = 0$). In that case the conventional rule of equating marginal damage to marginal abatement costs applies which is exactly the optimality condition in a model where the technology is described by an equation such as (1), where the materials flow conditions (2c), (2d) and (2h) are ignored, and where (7) satisfies $P_{r_{a1}} = P_{r_{a2}} = P_{r_c} = 0$. But if at least one of the other residuals also contributes to pollution ($P_{r_{a1}} > 0$ or $P_{r_{a2}} > 0$ or $P_{r_c} > 0$), an inequality sign applies in (15). To be more specific, suppose $P_{r_{a1}} > 0$ and/or $P_{r_{a2}} > 0$ and $P_{r_c} = 0$ which implies $MP_p \cdot P_e > \hat{Y}_e$ due to (15). From (5) we know that $\tilde{Y}_e = F_m M_e$ and hence $\tilde{Y}_{ee} = F_{mm} M_e (1 + F_m / A_o) < 0$. Therefore, it is now optimal to abate *y*-residuals to the point where the marginal costs of abating *y*-residuals is (still) smaller than the marginal benefit from reducing *y*-emissions. In other words, it is now optimal to 'under-internalize' the externality caused by *y*-emissions because stepping up abatement increases the generation of abatement residuals that are assumed to contribute to

pollution. If however $P_{r_{a_1}} = P_{r_{a_2}} = 0$ and $P_{r_c} > 0$ and hence $MP_p \cdot P_e < \hat{Y}_e$, 'over-internalization' of the y-emissions externality is obviously optimal because increased abatement has the additional benefit of reducing post-consumption residuals that cause pollution by assumption. It is worth noting that although the inequality $MP_p \cdot P_e \neq \hat{Y}_e$ has the flavor of a second best outcome the contrary is true: unless $P_{r_{a_1}} = P_{r_{a_2}} = P_{r_c} = 0$ holds the marginal-damage-equals-marginal-abatement-cost rule fails to implement an efficient allocation, in general. Observe also, that owing to (12c) – (12e) optimality requires, in general, to deviate from the well-known marginal-productivity-equals-factor-(shadow)-price rule.

Recall that in the preceding discussion we *assumed* abatement of y-residuals ($a > 0$) to be optimal (footnote 19). It is an interesting question to ask whether a solution of (11) is feasible with $\lambda_m > 0$ and therefore $m_a = a = 0$. Suppose for a moment (which has been shown in Section 3.3 to be wrong) that $\hat{Y}_e = 0$ for all e satisfying (6). By making appropriate use of the Kuhn-Tucker conditions we can then readily show that zero abatement must always be suboptimal. Yet with $\hat{Y}_e > 0$ for all boundary points ($a = 0$) it can be proved by careful inspection of the Kuhn-Tucker conditions that *zero abatement may be optimal*. This outcome is the more likely, *ceteris paribus*, the greater is the marginal pollution from the other residuals.

4.3 Production-cum-recycling and pollution

Besides abatement, recycling is another process of materials transformation that is of great relevance for residuals management. We therefore extend our analysis now to model recycling with our main attention focussed on the implications of material balance accounting. More specifically, drawing on Eichner and Pethig (2001a, 2001b) we consider the production technology

$$y = F(\ell_y, m, n) \quad (18a) \quad r_c = y \quad (18f)$$

$$v = V(\ell_v) \quad (18b) \quad y = m + n_y \quad (18g)$$

$$s = S(\ell_s, q, r_{cs}) \quad (18c) \quad r_y = n - n_y \quad (18h)$$

$$q := m / y \quad (18d) \quad s + e = r_{cs} \quad (18i)$$

$$m = v + s \quad (18e) \quad r_{cs} = r_c \quad (18j)$$

(18a) describes the production of the consumer good by means of labor, ℓ_y , and two different kinds of materials, called m -material and n -material. (Recall that in (2) we dealt with one kind of material only). The m -material is completely embodied in output y along the amount n_y of the n -material, (18g). The difference, $n - n_y$, is an y -residual²⁰, (18h). As a consequence, the consumer good consists of a material mix conveniently measured by $q \in]0, 1[$, (18d), which represents the content of m -material per unit of the consumer good. For short, we will refer to q as *material content*. (18b) describes the extraction of m -material (with v for virgin material) using labor as an input. (18c) specifies the recovery of m -material, s , (with s for secondary material) from consumption residuals, r_{cs} , and labor, ℓ_s . Virgin and secondary m -material is homogenous, (18e). (18e) – (18i) serve to satisfy the MBP. A quick consistency check of (18) reveals that $r_y = n - n_y$ and $e = r_{cs} - s = r_c - s = y - s = qy + (1 - q)y - s = m + n_y - s = v + n_y$. Hence in total, the amounts n and v of the n - and m -materials, respectively, are ultimately discharged into the environment. The ‘throughput’ of m -materials turns out to be equal to the amount of virgin m -material, v , and v is the smaller, *ceteris paribus*, the more m -material is recovered in the recycling process. The model is thus shown to keep track of both kinds of materials ‘from cradle to grave’.

The inclusion of the variable q (material content) as an argument in the recycling function, S , is necessary for correct materials balance accounting. To see that, consider some quantities r_{cs}^o, ℓ_s^o, s^o and $q^o \in]0, 1[$ such that $s^o = S(\ell_s^o, q^o, r_{cs}^o)$ and $q^o r_{cs}^o = s^o$. If we would set $S_q \equiv 0$ for all q , we would have $s^o = S(\ell_s^o, q, r_{cs}^o)$ for all $q < q^o$. But choosing some $q^1 < q^o$ implies $s^1 := q^1 r_{cs}^o < s^o = q^o r_{cs}^o$. Hence the production technology (18c) violates the law of mass conservation unless $S_q > 0$ holds.²¹ With slight modifications to our procedure in Section 4.2 (notably the endogenization of labor) we now complete the model by adding the (self-explanatory) equations

$$u_i = U^i(\underset{-}{\ell_i}, \underset{-}{p}, \underset{+}{y_i}), \quad p = P(\underset{+}{e}, \underset{+}{r_y}), \quad y \geq \sum_i y_i, \quad \sum_i \ell_i \geq \ell_s + \ell_v + \ell_y, \quad \bar{n} \geq n. \quad (19)$$

An efficient allocation of the recycling economy (18) and (19) can be characterized by solving the associated Lagrangean defined analogous to (11) in Section 4.2. If an interior solution

²⁰ Setting $r_y \equiv 0$ would make both materials perfect substitutes. Apart from the limiting case of indifference production efficiency would then imply to use one of them only.

²¹ Quite obviously, raising q facilitates recycling. Fullerton and Wu (1998) modeled a similar idea by introducing a variable into the recycling function for the ease of disassembling residuals or for ‘recyclability’ as they term it.

to this Lagrangean exists^{22, 23}, we show by using standard arguments that the solution satisfies²⁴

$$\sum_i \frac{U_p^i P_e}{U_\ell^i} = \frac{1}{S_\ell} - \frac{1}{V_\ell} > 0, \quad (20)$$

$$\sum_i \frac{U_p^i P_{r_y}}{U_\ell^i} + \frac{S_q}{y S_\ell} = \frac{F_m}{F_\ell} - \frac{1}{V_\ell} > 0. \quad (21)$$

According to (20) it is optimal to reduce recycling residuals, e , to the level where the induced marginal environmental benefit (right side of (20)) equals the difference between marginal labor cost of secondary material and marginal labor cost of virgin material. This difference is a wedge between marginal labor productivities in two different production processes for generating one and the same homogeneous product. When looking at this wedge in isolation it constitutes an inefficiency in the allocation of labor, since we can increase aggregate material output, $v + s$, quite obviously, by shifting some labor input from secondary material production to virgin material production. The wedge on the right side of (20) represents the marginal costs of reducing recycling materials caused by ‘over-expanding’ the production of secondary material. Likewise, the right side of (21) constitutes the marginal costs of reducing production residuals. These marginal costs are equal to the difference between marginal labor costs of m -material in the production of y , F_m / F_ℓ , and marginal labor costs of m -material extraction, $1/V_\ell$. Like the right side of (20) discussed above, this difference also consists of a ‘production distortion’ or wedge. The left side of (21) contains the associated total marginal benefit of reducing production residuals: the marginal environmental damage avoided plus the value of an increase in production of secondary material induced by a small increase in material content, q .

Consider now an unregulated market economy where producers and recyclers are independent profit maximizers and suppose, all agents face a uniform wage rate as well as a uniform price for virgin and secondary m -material. There is no direct or indirect market for material content.

However, in contrast to the ‘material content’ in our present model, the ‘recyclability’ in Fullerton and Wu (1998) is not related to the materials flow at all.

²² We note in passing that the inclusion of material content, q , as an endogenous variable renders the constrained maximization problem non-concave and hence jeopardizes the existence of a solution. This is not a trivial, purely technical observation, since after all, it was necessary to include q in order to satisfy the MBP.

²³ Analogous to our previous discussion on abatement, a boundary solution with zero recycling may be optimal. This scenario applies when the environmental damage caused by releasing recycling residuals is sufficiently great as compared to the damage caused by releasing consumption residuals (on the assumption that consumption residuals are directly discharged rather than used as recycling inputs).

Straightforward calculations show that the associated general competitive equilibrium exhibits $F_m / F_\ell = 1/V_\ell$ and $S_\ell = V_\ell$, which is an obvious violation of (20) and (21). When we assume $P_e = P_{r_y} \equiv 0$, (20) is not violated anymore in equilibrium but (21) still is. In other words, apart from the pollution externalities that emerge when $P_{r_y} > 0$ or $P_e > 0$ there is an additional efficiency-reducing externality at work caused by the material content, q . Since q is an attribute of the consumer good determined by its producer, we refer to that externality as *product design externality*. In choosing the material content of the consumer good the producer is led by minimization of his or her private costs without paying any attention to the needs or wants of the recycler. But the material content of the consumer good as fixed by the producer is an attribute of the consumption residuals, too, and hence also an attribute of the recycler's residuals input, r_{cs} . Although the material content of r_{cs} is not among the recycler's decision variables, it has an impact on the productivity of recycling through $S_g > 0$, nonetheless. It follows that a socially optimal level of q must be a 'green product design' that compromises between the interests of the producer and the recycler.

To sum up, as an implication of describing the materials flow correctly we found that upstream decisions on the quantities, the composition and transformation of materials have downstream consequences not all of which competitive markets cope with in an efficient way. Moreover, the identification of the product design externality demonstrates that besides pollution externalities other externalities may also be linked to the materials flow, when the MBP is duly taken into account.

5. Did environmental economists learn A&K's lessons?

With a synoptic view on our preceding discussion and with A&K's analysis as a base line we now summarize what we believe are the three principal messages of the materials flow approach to pollution and add a few comments on whether and how subsequent writing in environmental economics dealt with this research program launched by A&K some thirty years ago.

²⁴ Recall that we set all weight coefficients equal to unity. This is the reason why (20) and (21) appear to be inconsistent in dimensions, at first glance.

(I): *The concept of externality is an appropriate and powerful device to capture the economy-environment interdependence created by the discharge of residuals, the resulting environmental degradation and its feedback effects on the economy.*

The question whether the profession learnt this lesson has been answered already in the affirmative in Section 2. A&K argued convincingly that the crucial link between residuals discharge and pollution externalities is the assimilative capacity of the environment that depends on the kind of residuals, the kind of environmental medium used as a receptor of residuals, its location, and on the load of past residuals emissions already accumulated in that medium. Yet their ‘linear’ approach provided only limited insight in the complexity of the essentially ‘non-linear’ relationship between residuals and externalities. Therefore, it doesn’t seem unfair to say that as far as their formal model is concerned, A&K didn’t provide a completely satisfactory foundation for their sound and basic message that pollution *externalities* are *normal, inherent, inevitable and pervasive*. But invoking the law of mass conservation as an honorable, non-partisan and unsuspected witness, they certainly convinced the profession that it is the *unwanted residuals* that are *normal, inherent, inevitable and pervasive*. Subsequent conceptual and applied work greatly improved our understanding of the interface between the economy and the environment. Yet it seems to us that the direction taken by the profession was essentially to elaborate on the research program envisioned by A&K.

Looking further ahead in the future a slight caveat may be in order regarding the role of the externality concept. Although we think the externality perspective is still the backbone of environmental and resource economics, some contributors to the analysis of long-term effects and sustainability challenge this view and call for restrictions in the use of environmental services by invoking a precautionary principle that ought to be adhered to irrespective of whether an externality is to be internalized or not. This ongoing discussion may have the potential to change our perception of the central role of externalities. But a detailed discussion of this controversial and complex issue is beyond the scope of the present paper.

(II): *The ‘materials balance approach to pollution’ requires to analyze joint production and externalities in a general equilibrium framework.*

Since residuals are inherent, significant in size and pervasive in all economic sectors, residuals analysis and management affects all economic sectors and triggers repercussions in many sectors. Likewise, the feedback effects of environmental degradation on the economy have an impact on many economic agents and sectors. A&K deserve the merit for having made that point forcefully by presenting their analysis in a general equilibrium framework.

In our view, such a general equilibrium analysis is an adequate and necessary method for materials flow analysis. This thesis may come as a surprise, at first glance, because the notion of materials flow has a dynamic (time) dimension to it while the equilibrium as discussed by A&K and in the present paper is inherently static. But note that within one and the same period, which may even be a point in time, the activities of extraction, production, recycling/abatement, disposal and the transactions associated with these activities are all understood to occur in a sequence of time. It is therefore appropriate to refer to ‘upstream’ or ‘downstream’ activities in static general equilibrium analysis thus indicating that a materials flow perspective is adopted (e.g. Vatn 1998, Walls and Palmer 2000).

It didn’t take great efforts to convince the academic community of the need to study large-scale pollution problems in a general equilibrium framework. Interdependent activities always have been a central focus of economics, and the relevance of interdependencies in large-scale residuals and pollution issues was quite obvious. Following the A&K lead, the early and influential textbook of Baumol and Oates (1975) warned against the fallacies of partial equilibrium analysis in environmental economics.²⁵ It is appropriate to add, though, that there is still much room and a need for informative partial equilibrium studies, too, such as the work on the household’s waste decisions reviewed by Kinnaman and Fullerton (2000).

Following A&K, numerous subsequent studies applied general equilibrium analysis with economy-environment interaction ranging from highly aggregated one-sector approaches to models with multiple sectors. The last decade saw a strong shift toward computable general equilibrium analysis as an important tool for policy advice focusing on various kinds of issues with significant interdependence such as marketable permits, eco-taxes and climate change (Conrad 2002).

Attesting the profession to have learnt the lesson of studying the materials flow and the externalities related to it in a general equilibrium framework does not yet imply that the materials flow is studied in line with the MBP.

(III): *The description of materials flow is incorrect, if it is not in line with the MBP.*

We don’t believe any environmental and resource economist needs to learn the lesson that the law of mass conservation is an ‘eternal truth’ which is at nobody’s disposal. But this insight did not prevent most of them from analyzing environmental issues with blatant disregard of

the MBP: “The relevance of the first law [of thermodynamics, R.P.] is widely accepted in principle among environmental and resource economists (see Ayres and Kneese, 1969), though the extent to which material balance is considered in practice in neoclassical analyses is more limited.” (Pezzey and Toman 2002, 202). On average, the professional interest in and the awareness of the need to correctly describe the materials flow remained moderate.²⁶ We are therefore led to conclude that the profession was not willing to follow the path of virtue of modeling the materials flow in line with the MBP so dearly recommended to them by A&K. In many studies, production processes are still viewed like 30 years ago, namely “... in a manner that is somewhat at variance with the law of conservation of mass”. (A&K 1969, 283).

This observation begs two questions: What are the reasons for that refusal? What are the consequences?

One of the reasons for the disregard of the MBP may be to avoid the vast increase in analytical complexity that arises when technologies are assumed to be non-linear and the materials flow is described correctly. The difference in complexity in applying the technologies (1) or (2) illustrates that point well. Yet economists are not renowned for avoiding analytical complexity when they consider it important. Therefore, laziness doesn’t seem to be a convincing explanation for their reluctance. Another reason for the use of truncated technologies such as (1) is simplification in conceptual and/or heuristic analyses as a means, similar as the *ceteris-paribus* clause, to restrict and thus sharpen the focus on a specific issue. Though this is an acceptable argument, it is less understandable that most authors do not even mention the incompleteness of the materials flow they focus on. Moreover, CGE studies aiming at consulting policy makers certainly do not (or should not) qualify as purely expository. Yet Conrad’s (2002) survey on such studies does not even mention the MBP as an essential ingredient. CGE-studies dealing with abatement apply the emission-tax-equals-marginal-abatement-costs rule and hence ignore the MBP-induced deviations from that rule discussed in Section 4.2.

Since all other reasons for the profession’s reluctance to employ the MBP that we are aware of are directly related to the consequences of disregarding the MBP, we now turn to looking at the consequences. To be more specific we reconsider the production-cum-abatement model that was extensively discussed in Sections 3.3 and 4.2. The question then is what are the con-

²⁵ A more recent stunning example for plausible but misleading partial equilibrium arguments is the double dividend debate in the context of revenue-neutral ecological tax reforms when distortionary taxes preexist (Bovenberg and de Mooij 1994).

²⁶ It is interesting to observe, however, that during the last years this topic received growing attention again in a number of studies with strong links to physics (e.g. Baumgärtner 2000, 2002, Faber et al. 1998).

sequences of employing the ‘truncated’ technology (1) instead of the correct production technology (2)? In view of our analysis above, the answer is straightforward:

Applying the truncated technology (2) amounts to focusing on a single type of residuals (y -residuals) only which is generated in a production activity that inevitably creates abatement residuals and post-consumption residuals in addition to y -residuals. Piece-meal, individual-pollutant approaches turn out to be inefficient *if the residuals ignored also cause pollution*. As we demonstrated in Section 3.3, incompletely described production technologies ignore interdependencies among wanted and unwanted outputs. If any of the ignored residuals contributes to pollution, the resulting rules for optimal levels of emissions (of all residuals, ignored and not ignored ones) fail to induce an efficient allocation and so does environmental management guided by those rules. The bias from basing policy recommendations on (1) rather than on (2) is the greater the larger are the contributions to pollution of abatement and post-consumption residuals. Therefore, an *integrated, comprehensive* environmental policy is the only policy approach capable to restore efficiency. In conclusion, the consequences of disregarding the MBP are grave in the preceding scenario.

Although (2) is a truncated technology, it leads to correct (policy) conclusions, if and only if (i) y -residuals are detrimental to the environment and (ii) all other residuals (abatement residuals and post-consumption residuals) contribute to pollution in a negligible way only. Hence (1) can be ‘reconciled’ with the MBP if waste abatement as implied by (1) really means the transformation of detrimental y -residuals in utterly harmless abatement residuals. Using (1) can also be interpreted as employing a partial free disposal assumption: If this assumption turns out to be correct, i.e. if all residuals other than e can be disposed of without any social cost, then it suffices to employ (1) rather than (2). For studying empirical pollution problems, partial materials flow analysis may be an acceptable approximation in some cases but will be grossly inadequate in others. Under which conditions a comprehensive materials flow analysis is needed always depends on the problem at hand.

The preceding observations show that there are good reasons to take a more pragmatic view on the issue of incompletely described materials flows keeping in mind the environmental and resource economists’ primary interest in the links between residuals, environmental degradation and the associated externalities. Since not all residuals are pollutants and not all pollutants cause equally severe damages it follows that focusing on pollutants is more important than focusing on *all* residuals and focusing on heavy-damage pollutants is more important than focusing on *all* pollutants. The information that all mass taken from nature will eventu-

ally be returned to nature is of limited value since environmental damage caused by the discharge of residuals depends on the physical and chemical attributes of those residuals, on the location of discharge and on the assimilative capacities of the recipient environmental media.

Model building has always been about focusing attention on - and restricting it to - what are believed to be the important building blocks and interdependencies. These choices are subjective, to a large extent, and so are decisions of the model builder as to which parts and aspects of the materials flow to include in and which to exclude from his or her analysis. It is as legitimate to criticize incomplete mappings of the materials flow as it is to criticize any (other) 'restrictive' or 'simplifying' assumption. But if the MBP with its implied linear constraints is not essential for the problem to be tackled, the analysis need not be burdened with tedious mass balance accounting. To follow the maxim of Okham's razor is a reasonable research strategy in economics at large and therefore also when residuals and pollution are at issue.

Making the case for keeping the MBP at a low profile with an appeal to pragmatism doesn't amount to a general absolution of economists from the obligation to describe economic activities in line with the law of mass conservation. It leads us, however, to reject the sweeping argument that all those pieces of environmental-economic research are fundamentally flawed which are found guilty of not properly regarding the MBP. 'Economics bashing' per se is not a satisfactory substitute for constructive alternative analysis (Kaufmann 2001). The MBP needs and ought to be adhered to whenever it is an essential component of substantive analysis. Yet invoking the 'spirit of the MBP' remains unsatisfactory in a discourse that exhausts itself in simply *defining* a mass balance framework with more or less sophisticated terms and identities. Models that consist of a set of accounting identities only are useful as a frame of reference, "... but inadequate when it comes to explaining the working of the economy" (Mäler 1974, p. 7). The test of the usefulness of studying the materials flow with full regard of the MBP are substantive non-trivial results both in theoretical and applied work obtained through the application of the mass balance concept.

References

- Ayres, B. (1972), "A materials-process-product model", in: Kneese, A.V., and Bower, B.T. (eds.), *Environmental Quality Analysis. Theory and Method in Social Sciences*, Baltimore
- Ayres, B., and Kneese, A. (1969), "Production, consumption and externalities", *American Economic Review* 59, 282-97
- Baumgärtner, S. (2002), "Thermodynamics and waste generation", in: K. Bisson and J. L. R. Proops (eds.), *Waste in Ecological Economics*, Edward Elgar, Cheltenham, Northampton
- Baumgärtner, S. (2000), *Ambivalent Joint Production and the Natural Environment. An Economic and Thermodynamic Analysis*, Physica-Verlag, Heidelberg, New York

- Baumgärtner, S., and Jost, F. (2000), “Joint production, externalities, and the regulation of network externalities”, *Environmental and Resource Economics* 16, 229-251
- Baumol, W. J., and Oates, W. E. (1975), *The Theory of Environmental Policy*, (2nd edition 1988), Englewood Cliffs, N.J.
- Bergstrom, T. C. (1976), “How to discard free disposal – at no cost”, *Journal of Mathematical Economics* 3, 131-134
- Bovenberg, A. L., and de Mooij, R. A. (1994), “Environmental levies and distortionary taxation”, *American Economic Review* 94, 1085-1089
- Conrad, K. (2002), “Computable general equilibrium models in environmental and resource economics”, in: T. Tietenberg and H. Folmer (eds.), *The International Yearbook of Environmental and Resource Economics 2002/2003*, Edward Elgar, Cheltenham, Northampton, 66-114
- Dasgupta, P. (1982), *The Control of Resources*, Basil Blackwell, Oxford
- Debreu, G. (1959), *Theory of Value. An Axiomatic Analysis of Economic Equilibrium*, New York
- Eichner, T., and Pethig, R. (2001a), “Recycling, producer responsibility and centralized waste management”, *Finanzarchiv* 57, 333-360.
- Eichner, T., and Pethig, R. (2001b), “Product design and efficient management of recycling and waste treatment”, *Journal of Environmental Economics and Management* 41, 109-134.
- Faber, M., Proops, J., Speck, S., and Jost, F. (1999), *Capital and Time in Ecological Economics: Neo-Austrian Modelling*, Edward Elgar, Cheltenham, Northampton
- Faber, M., Proops, J. L. R., and Baumgärtner, S. (1998), “All production is joint production – a thermodynamic analysis”, in: S. Faucheux, J. Gowdy and I. Nicolai (eds.), *Sustainability and Firms: Technical Change and the Changing Regulatory Environment*, Edward Elgar, Cheltenham, Northampton, 131-158
- Fullerton, D., and Wu, W. (1998), “Policies for green design”, *Journal of Environmental Economics and Management* 36, 131-148
- Kapp, W. (1972), “Environmental disruption and social costs: A challenge to economics”, in: OECD (ed.), *Political Economy of Environment. Problem of Methods*, Paris
- Kaufmann, R. (2001), “The environment and economic well-being”, in: H. Folmer, H. L. Gabel, S. Gerking and A. Rose (eds.), *Frontiers in Environmental Economics*, Edward Elgar, Cheltenham, 36-56
- Kinnaman, T. C., and Fullerton, D. (2000), “The economics of residential solid waste management”, in: T. Tietenberg and H. Folmer (eds.), *The International Yearbook of Environmental and Resource Economics 2000/2001*, Edward Elgar, Cheltenham, Northampton, 100-147
- Kneese, A. V., Ayres R. U., and d’Arge, R. C. (1970), *Economics and the Environment. A Material Balance Approach*, Washington
- Mäler, K.G. (1974), *Environmental Economics: A Theoretical Inquiry*, Baltimore
- Perrings, C. (1986), “Conservation of mass and instability in a dynamic economy-environment system”, *Journal of Environmental Economics and Management* 13, 199 – 211

- Perrings, C. (2001), "Modelling sustainable ecological-economic development", in: H. Folmer, and T. Tietenberg (eds.), *The International Yearbook of Environmental and Resource Economics 2001/2002*, Edward Elgar, Cheltenham, Northampton, 179-201
- Pethig, R. (2001), "On the future of environmental economics", in: H. Folmer, H. L. Gabel, S. Gerking and A. Rose (eds.), *Frontiers of Environmental Economics*, Edward Elgar, Cheltenham, Northampton
- Pethig, R. (1979), "Environmental management in general equilibrium. A new incentive compatible approach", *International Economic Review* 20, 1-27
- Pethig, R. (1976), "Pollution, welfare and environmental policy in the theory of comparative advantage", *Journal of Environmental Economics and Management* 2, 160-169
- Pezzey, J.C.V., and Toman, M.A. (2002), "Progress and problems in the economics of sustainability", in: T. Tietenberg and H. Folmer (eds.), *The International Yearbook of Environmental and Resource Economics 2002/2003*, Edward Elgar, Cheltenham, Northampton, 165-232
- Russell, C.S., and Spofford, W.O. (1972), "A quantitative framework for residuals management decisions," in: A. V. Kneese and B. T. Bower (eds.), *Environmental Quality Analysis. Theory and Method in Social Sciences*, Baltimore
- Sen, A. G. (1985), "Goals, commitment, and identity", *Journal of Law, Economics and Organisation* 1, 341-355
- Stiglitz, J. E. (1974), "Growth and exhaustible natural resources: efficient and optimal growths paths", *Review of Economic Studies*, Symposium on the Economics of Exhaustible Resources, 123-137
- van den Bergh, J.C.J.M., and Nijkamp, P. (1994), "Dynamic macro modelling and materials balance: Economic-environmental integration for sustainable development", *Economic Modelling* 11, 283 - 307
- Vatn A. (1998), "Input versus emission taxes: Environmental taxes in a mass balance and transaction costs perspective", *Land Economics* 76, 413 - 429
- Walls, M., and Palmer, K. (2000), "Upstream pollution, downstream waste disposal, and the design of comprehensive environmental policies", *Journal of Environmental Economics and Management* 27, 1-15
- Weinberg, M., and Newbold, S.C., "Imagining externalities: Material balance and the environmental economics literature", this volume
- Tahvonen, O., and Kuuluvainen, J. (1993), "Economic growth, pollution, and renewable resources", *Journal of Environmental Economics and Management* 24, 101-118